

# Towards a Domain – Specific Comparative Analysis of Data Mining Tools

May Oo

Daniel Lozovikas

Ramesh Subramanian

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/ciima>



Part of the [Management Information Systems Commons](#)

---

## Towards a Domain – Specific Comparative Analysis of Data Mining Tools

**May Oo**

Quinnipiac University

**Daniel Lozovikas**

Quinnipiac University

**Ramesh Subramanian**

Quinnipiac University

Hamden, Connecticut, USA 06518

### ABSTRACT

*Advancement in technology has brought in widespread adoption and utilization of data mining tools. Successful implementation of data mining requires a careful assessment of the various data mining tools. Although several works have compared data mining tools based on usability, opensource, integrated data mining tools for statistical analysis, big/small scale, and data visualization, none of them has suggested the tools for various industry-sectors. This paper attempts to provide a comparative study of various data mining tools based on popularity and usage among various industry-sectors such as business, education, and healthcare. The factors used in the comparison are performance and scalability, data access, data preparation, data exploration and visualization, advanced modeling capabilities, programming language, operating system, interfaces, ease of use, and price/license. The following popular data mining tools are assessed: SAS Enterprise Miner, KNIME, and R for business, Moodle Learning Analytics, Blackboard Analytics, and Canvas for education, and RapidMiner, IBM Watson Health, and Tableau for healthcare. It also discusses the critical issues and challenges associated with the adoption of data mining tools. Furthermore, it suggests possible solutions to help various industries choose the best data mining tool that covers their respective data mining requirements.*

**Keywords:** Data mining tools, Industry-wide comparison, Business, Education, Healthcare, Comparative analysis

---

## INTRODUCTION

Data mining has emerged as a fundamental and indispensable practice in industries worldwide. It enables organizations to extract, analyze and leverage valuable insights from large amounts of data. As the volume and complexity of data grows, traditional methods of analysis have become inadequate and inefficient. In response to this challenge, data mining tools have emerged as powerful aids, empowering businesses to efficiently navigate complex data and make rational decisions. By leveraging advanced algorithms and computational techniques the data mining tools automate the process of extracting knowledge from large datasets. As a result, organizations in various industries such as business, education, and healthcare have turned to data mining as an important tool to gain competitive advantage and enhance their overall performance. This comparative study provides a deep understanding of popular data mining tools and unlocks their potential and usefulness of them across sectors within business, education, and healthcare.

### *Background*

At its core, data mining is based on an iterative process of data exploration, model building, and interpretation. The process begins by collecting and integrating large amounts of structured and unstructured information from multiple sources, including databases, websites, social media, sensors, etc. Once data is collected, preprocessing takes place. Preprocessing is the first step where data is properly formatted to ensure accuracy and quality. The next step is to select and apply appropriate data mining models and associated techniques to the prepared data set to identify patterns and relationships in the data. The techniques can be divided into broad categories such as classification, clustering, association rule mining, regression analysis, and anomaly detection. All these data mining techniques can help analyze different data from different perspectives. For example, classification algorithms are used to display categorical results or assign objects to pre-defined classes, while clustering algorithms collect similar data points based on their common characteristics. Association rule mining reveals correlations and dependencies between variables, while regression analysis tests statistical relationships between variables to determine statistics' internal effects.

Anomaly detection, on the other hand, identifies unusual patterns or outliers within the data (Chen, Han, and Yu, 1996).

Hence, it is important to select the appropriate techniques suitable for the data and the problem being addressed to obtain the optimal outcomes. The last step of the data mining process is data interpretation which refers to the process of using different data mining techniques to draw relevant conclusions for decision making (Linoff & Berry, 2011).

In this way, data mining helps organizations recognize industrial trends and how to respond to them, or identify hidden relationships within existing data. However, every industry uses data mining differently to benefit their organizations. Depending on the industry, it offers specific benefits. For example, data mining helps businesses detect fraudulent activities by locating suspicious behavior and false patterns. It can also predict future trends based on past behavior; however, this can be effective only if historical data are available (Hyperverge, 2022).

In education, data mining helps to improve quality of education by investigating the performance of instructors (Fujita, 2019) as well as predicting the performance of students (Agaoglu, 2016; Baker, 2010). This allows educators to improve their teaching skills and take precautions for students who are at risk (Ünal, 2020). In the healthcare industry, data mining helps healthcare insurers to detect fraud and abuse, healthcare organizations to make management decisions, providers to identify effective treatments and best practices, and patients to receive better and more affordable healthcare services (Koh & Tan, 2011).

Despite its many advantages, data mining faces many challenges and limitations. One of the biggest challenges is finding the right data mining tool that meets organizations' goals. Data mining tools are complex and require high levels of training to use them effectively. Other challenges include computational and storage requirements due to the volume and complexity of data, which requires organizations to invest in robust infrastructure and scalable solutions. Limitations of data mining in various industries are as follows. Limitations in business include the risk concerning information accuracy, privacy, ethical issues, and data security. To address this, businesses can maintain transparency in mining data and leverage privacy-enhancing technologies such as masking and encryption to protect sensitive data (Şimşek, 2023). Limitations in education include having enough educational data, misinterpretation of the data, and the need for smooth integration of learning management systems with other information system applications. Promoting information management and utilizing advanced and scientific statistical tools and mining expertise in the education sector will help to thoroughly examine data and boost management intelligence intensity (Iwatani, 2018; Romero, 2012; Zhang et al., 2022). Limitations in healthcare include the reliability of medical data, data sharing between organizations, and inaccurate predictions. Research suggests that more work is needed to show evidence of data mining's impact on healthcare decision-making, healthcare services, and patient care (Mantas, 2017).

### ***Prior Work***

Existing work in data mining tools is examined in this section. Previous studies on comparative study of data mining tools are mostly focused on several factors, including functionality, open- source availability, integration and statistical

analysis, scalability, data visualization (Bhinge, 2015; Rawat, 2017; Santos-Pereira, Gruenwald, and Bernardino, 2022; Slater et al., 2017). Moreover, data mining tools are compared based on dashboard availability, text mining availability, and deployment (standalone or network installations) (Saabith & Fareez, 2018). However, despite a growing body of literature on comparative study of data mining tools, none of them has suggested data mining tools for specific industry sectors. To the best of our knowledge, this is the first work that has compared popular data mining tools in business, education, and health care based on their critical capabilities.

### ***The Problem***

Data mining tools are indispensable in the present-day scenario. Yet, they present several challenges that organizations need to address for effective implementation. Firstly, these tools can be highly technical and intricate, requiring specialized technical skills and training to navigate efficiently (Abraham, 2019). The learning curve may be steep, necessitating dedicated resources for training. Secondly, selecting the right tool can be challenging as different tools have varying capabilities, functionalities, and compatibilities with different data sources and formats (Hill, 2020). Choosing the wrong tool can lead to suboptimal outcomes and inefficient data analysis. Additionally, data mining tools employ different methods and algorithms, requiring organizations to be sophisticated enough to understand the approaches and select the one that aligns best with their goals (Verma, 2020). Furthermore, the complexity and cost of analytics tools may discourage small businesses from adopting these practices. Limited resources and a lack of technical skills can prevent them from using data mining effectively. A lack of standardization between tools can also lead to integration and compatibility issues, as each tool may have its own unique interface, programming language, and operating system requirements (Kim, 2018).

Scalability and performance are important considerations as data volumes continue to grow. Organizations require data mining tools capable of handling large datasets efficiently and providing timely results. Data privacy and security concerns arise due to the sensitive nature of the information involved in data mining. Organizations must ensure that their chosen tools adhere to robust security measures and ethical practices (Huang, 2019).

Cost and budget constraints also play a role, as some data mining tools can be expensive, making them inaccessible to organizations with limited budgets (Schultz, 2021). The affordability of equipment and value for money should be carefully considered in the selection process. Additionally, successful implementation of data mining tools requires skilled professionals with expertise in data analysis, statistical modeling, and programming (Rana, 2022).

To overcome the shortage of skilled workers, organizations may need to invest in internal training programs or hire outside contractors. Considering ethical aspects is also extremely important in the process of data mining. This involves dealing with biases in data, respecting privacy, and avoiding any potential discrimination. Organizations need to be conscious of these ethical concerns and adopt responsible and transparent approaches while utilizing data mining tools (Joshi, 2019). Thus, while data mining tools offer significant benefits, organizations must navigate the challenges and ethical considerations associated with their implementation. By investing in training, selecting appropriate tools, addressing scalability, and cost restraints, and adopting responsible and transparent practices, organizations can harness the power of data mining while upholding ethical standards and promoting positive outcomes.

### *Approaches To Solutions*

In today's data-driven world, organizations face the challenge of harnessing the full potential of data mining. To overcome this challenge, it is crucial for organizations to carefully select the right data mining tools that align with their unique business need. Scalability and compatibility are important factors to consider (Liao Hu 2020), to ensure that chosen tools can effectively handle the volume and complexity of data. By tailoring data mining tools to specific requirements, organizations can greatly enhance their ability to extract meaningful insights. To do so, organizations may decide to acquire several tools, for example, installing a tool with dashboarding capabilities to keep managers informed on business matters, a full programming suite to capture and build data for its marketing arm, and an interrogation tool for auditors to identify fraud (Ramamohan et al., 2012).

Selecting tools that align with the nature of the domains, their data, and desired outcomes can improve the accuracy and relevance of analyses. For example, in the healthcare sector, advanced machine learning algorithms in data mining tools have enabled the identification of disease patterns and risk factors from extensive patient datasets (Huang 2021). Data mining tools also help with the prediction of diseases and improve diagnostic accuracy by eliminating prejudice and subjectivity (Durairaj & Ranjani, 2013).

In addition to selecting the right tools, organizations should prioritize training programs for employees so they can use data analysis tools effectively.

Advances in data analytics technology require professionals who can master these tools and use them to their full potential.

Training programs can bridge the gap between employees' existing knowledge and the need for data mining tools, empowering them to generate insights from complex data sets, according to (Katal 2020).

By investing in training and employee capacity building, organizations can increase their data mining capabilities to achieve better decision-making processes.

### ***The Promise of AI***

Data mining tools have advanced significantly, offering exciting opportunities for researchers in the future. One promising area of progress is the integration of artificial intelligence and machine learning techniques with these tools. Recent studies have demonstrated that incorporating AI algorithms, such as deep learning, can yield substantial benefits in the data mining process (Yang et al., 2020).

These techniques excel at handling complex and unstructured data, enabling organizations to gain deeper insights.

The integration of AI and machine learning algorithms in data mining brings numerous advantages. These algorithms can analyze vast amounts of data, uncovering patterns and relationships that may otherwise go unnoticed.

This is particularly valuable in today's digital landscape, where unstructured data, like text, images, and videos, are abundant. By leveraging AI techniques, organizations can discover valuable insights and make well-informed decisions based on comprehensive and accurate analyses (Yang et al., 2020).

Alongside AI integration, the best data mining solutions prioritize usability and accessibility. User-friendly interfaces, intuitive workflows, and visualizations play a crucial role in encouraging the adoption of these tools, especially among non-technical users (Stahl et al., 2022). It is vital to empower domain experts and decision-makers who may lack extensive technical knowledge or programming skills. Intuitive interfaces and workflows make it easier for these users to interact directly with the tools, explore data, and derive actionable insights.

Visualizations, such as charts, graphs, and interactive dashboards, facilitate the interpretation and communication of complex information, enabling users to make informed decisions (Stahl et al., 2022). By carefully selecting data mining tools based on business characteristics, providing comprehensive training programs, and considering prospects like AI integration, improved usability, organizations can significantly enhance their data mining capabilities and unlock the full potential of their data.

### ***The Importance of Adopting the Correct Tools***

There are several challenges in adopting data mining tools. One of the key challenges is getting senior management to understand its potential, finding enough talent to use the tools, and creating the right data fabric to manage data in a unified environment. It is also crucial to get senior management buy-in.

Senior management often is hesitant about investing for scale, such as data mining tools and training. To make matters worse, frontline managers also frequently lack confidence that data mining tools will improve their decision-making because they do not understand the suggestions or recommendations from the data mining tools. Moreover, they believe data mining tools are not user-friendly or are not going to embed into established workflows and processes. Finally, an organization's core processes can be a barrier as management-approval processes have not kept up with the technology advancements, leading to the risk of falling behind (Gordon, Perrey, and Spillecke, 2015).

The above challenges can be mitigated by having stakeholder engagement, resource allocation, training programs, frequent tool updates, and *knowledge of popular data mining tools*. Having stakeholder engagement during the acquisition process of data mining tools will help organizations to scale up and make an impact. Resource allocation will allow the organizations to use their budget on areas that can most significantly impact their growth. Training in data mining tools will help employees become better with the tools and overcome performance gaps due to lack of knowledge or skills. Frequent tool updates will explain why the suggestions or recommendations were being made. *Knowledge of popular data mining tools will help organizations select the best data mining tool before making an investment in any proprietary solution*. In order to help this, the following comparative study was conducted.

## A COMPARISON OF DATA MINING TOOLS IN VARIOUS INDUSTRIES

Data mining tools are software packages that have the ability to analyze large amounts of data to discover meaningful patterns and predict outcomes (Tan et al., 2006). Several data mining tools are available in the market for information extraction and data analysis. In this comparative study, several data mining tools (SAS Enterprise Miner, KNIME, and R for business, Moodle Learning Analytics, Blackboard Analytics, and Canvas for education, and RapidMiner, IBM Watson Health, and Tableau for healthcare) have been selected based on popularity among various industries. The selected tools in Figure 1 are compared based on critical capabilities – performance and scalability, data access, data preparation, data exploration and visualization, advanced modeling, programming language, operating system, interfaces, ease of use, and price/license (Table 1-3).

These critical capabilities are based on Gartner's report on advanced analytics platforms (Linden et al., 2016) and data science platforms (Linden et al., 2017) and defined as follows.



Performance and scalability: This refers to the time taken to load the data, create, validate and deploy the models.

Data access: This addresses the ability of a tool to access and integrate data from various sources and of different types (numeric, text, images, etc.)

Data preparation: This includes the ability to clean, transform, and filter data to prepare it for modeling.

Data exploration and visualization: This allows a range of exploratory steps, including interactive visualizations, to support data mining methods.

Advanced modeling: This provides the ability to create models that predict study outcomes or behaviors.

Programming language: This is the language used in data mining (SAS, R, Python, etc.)

Operating system: This refers to the compatibility with various operating systems (Windows, Mac, Linux, etc.)

Interfaces: This denotes the type of interface that a tool has (Graphical user interface, programmatic interface, command-line interface, etc.)

Ease of use: This refers to the ease of use of a tool.

Price/license: This provides the information on pricing and license of the tools.

In the next section, three examples of common data mining tools for each industry sector are presented.



**Figure 1. Data Mining Tools Used in Various Industries**

### *Data Mining Tools: General Business*

In business, data mining tools play an important role in gaining valuable insights from decision making and improving decision-making processes. Typically, the following data mining tools are used.

#### *SAS Enterprise Miner*

SAS Enterprise Miner provides advanced data and predictive modeling capabilities. It provides advanced data source connectivity, data cleaning, and transformation features. It supports graphical data exploration and visualization, as well as a wide variety of modeling techniques and algorithms. However, it has a steep learning curve and requires programming skills. Pricing for SAS Enterprise Miner is factory-based (SAS Enterprise Miner, 2023).

#### *KNIME:*

KNIME is an open-source data mining tool with extensive data mining and analysis capabilities. It supports data sources and formats, provides preprocessing and cleaning tools for storing data.

KNIME offers a wide range of machine learning and predictive models with rich visualization and dashboarding capabilities. It has an easy-to-use interface with drag-and-drop functionality. KNIME is free to use, and professional support is available (Knime, 2023). **R:** R is a programming language widely used for statistical analysis and modeling. It provides comprehensive statistical modeling capabilities and supports a variety of data sources and formats. R provides libraries and packages for visualization and data mining. However, effective use of the tool requires organizational and statistical skills. R is open-source and free to use (R, 2023).

	<b>SAS Enterprise Miner</b>	<b>KNIME</b>	<b>R</b>
Key Features	Advanced data mining and predictive modeling	Extensive data mining and analytics capabilities	Comprehensive statistical analysis and modeling
Performance and Scalability	High-performance analytics capabilities	Scalable processing for large datasets	Scalable and efficient processing
Data Access	Comprehensive data source connectivity	Support for various data sources and formats	Support for various data sources and formats

Data Preparation	Comprehensive data cleaning and transformation	Preprocessing and cleansing tools for data preparation	Data preprocessing and transformation capabilities
Data Exploration and Visualization	Graphical data exploration and visualization	Rich visualization and dashboarding capabilities	Visualization libraries and packages available
Advanced Modelling	Advanced modeling techniques and algorithms	Wide range of machine learning and predictive modeling algorithms	Extensive statistical modeling capabilities
Programming Language	SAS programming language for modeling and analysis	Utilizes various programming languages for analysis	R programming language for statistical analysis
Operating System	Windows and UNIX-based systems	Compatible with various operating systems	Compatible with various operating systems

Interfaces	Graphic user interface (GUI) and programmatic interface (IDE)	GUI and workflow editor	Command-line interface (CLI) and IDE
Ease of use	Steep learning curve, requires programming knowledge	User-friendly interface with drag-and-drop functionality	Requires programming and statistical knowledge
Price/License	Enterprise pricing, contact vendor for details	Open-source and free to use, with commercial support available	Open-source and free to use

**Table 1. shows the comparison of SAS Enterprise Miner, KNIME, and R based on their critical capabilities**

### *Data Mining Tools: Education*

In the education industry, data mining or learning analytics tools help to analyze educational data and improve learning outcomes. Typically, the following data mining tools are used.

***Moodle Learning Analysis:*** Moodle Learning Analysis focuses on delivering learning analysis for a Moodle-based learning management system. It provides interactive visualization and analysis tools for educational data, supports multiple programming languages, and provides userfriendly interfaces. Moodle Learning Analytics is open source and free to use (Moodle, 2023).

***Blackboard Analytics:*** Blackboard Analytics provides analytics for Blackboard learning management systems. It offers data cleansing and normalization features, rich visualization and dashboarding capabilities, and advanced analytics and predictive modeling. Blackboard Analytics has an easy-to-use interface with guided analytics, making it easy to use. Pricing depends on the application and specific solution (Blackboard, 2023).

**Canvas:** Canvas is a data mining tool specifically designed for Canvas-based learning management systems. It provides robust learning analytics capabilities, enabling educational institutions to analyze educational data and gain valuable insights to improve learning outcomes. Canvas supports a variety of data sources and formats, provides seamless integration with the Canvas course management system, and provides easily accessible data. The tool also offers data preparation features such as data preprocessing and cleansing, to ensure data is clean and ready for analysis.

With Canvas, users can use interactive visualization and search tools to explore tutorial data in depth, as well as advanced modeling techniques for predictive analytics. Supports multiple programming languages, has flexible interfaces for its usage, and offers pricing based on specific applications and solutions (Canvas, 2023).

Table 2 shows the comparison of **Moodle Learning Analytics, Blackboard Analytics, and Canvas** based on their critical capabilities.

	<b>Moodle Learning Analytics</b>	<b>Blackboard Analytics</b>	<b>Canvas</b>
Key Features	Learning analytics for Moodle-based learning management systems	Analytics for Blackboard learning management systems	Learning analytics for Canvas-based learning management systems
Performance and Scalability	Scalable analytics for large educational datasets	Scalable infrastructure for analyzing educational data	Scalable analytics for Canvas data
Data Access	Integration with Moodle learning management system	Integration with Blackboard learning management system	Integration with Canvas learning management system

Data Preparation	Preprocessing and cleansing tools for educational data	Data cleansing and normalization features	Data preprocessing capabilities
Data Exploration and Visualization	Interactive visualization and exploration tools for educational data	Rich visualization and dashboarding capabilities	Data exploration and visualization features
Advanced Modelling	Machine learning models for educational data analysis	Advanced analytics and predictive modeling	Advanced modeling and predictive analytics
Programming Language	Supports multiple programming languages	Utilizes various programming languages for	Supports multiple programming
		analysis	languages
Operating System	Compatible with various operating systems	Compatible with various operating systems	Compatible with various operating systems

Interfaces	Web-based interface and Application programming interface (API)	Web-based interface and API	Web-based interface and API
Ease of use	User-friendly interface with intuitive features	User-friendly interface with guided analytics	User-friendly interface with intuitive features
Price/License	Open-source and free to use	Pricing based on usage and specific solutions	Pricing based on usage and specific solutions

**Table 2. Comparison of Data Mining Tools in Education**

### ***Data Mining Tools: Healthcare***

In the healthcare industry, data mining tools help analyze patient data and support medical decision-making. Typically, the following data mining tools are used.

***RapidMiner:*** RapidMiner provides powerful data mining capabilities for healthcare analytics. It offers a wide range of preprocessing and cleaning tools, interactive visual search tools, and machine learning algorithms. RapidMiner supports multiple programming languages and has a user-friendly interface with drag-and-drop functionality. RapidMiner’s pricing options include free community membership (RapidMiner, 2023).

***IBM Watson Health:*** IBM Watson Health uses AI-powered analytics and insights to analyze healthcare data. It provides integration with healthcare data management, data cleansing and normalization features, and rich visualization and dashboarding capabilities. IBM Watson Health uses programming languages for analytics and provides web-based interfaces and APIs. Pricing depends on the application and specific solution (IBM, 2023).

***Tableau:*** Tableau is a popular data analysis tool widely used in the healthcare industry to analyze patient data and facilitate data-driven decision-making. With Tableau, healthcare professionals can leverage its advanced data visualization and analytics capabilities to gain valuable insights from healthcare data. Tableau provides integration with healthcare data management, allowing easy access to patient data for analysis. The tool provides features for data preprocessing,



cleansing, and normalization, ensuring data integrity and readiness for analysis. Tableau excels in providing rich visualization and dashboarding capabilities, allowing healthcare professionals to visually analyze and present data clearly and easily. It supports programming languages for analysis and provides web-based interfaces and APIs for easy integration with other systems. In terms of usability, Tableau offers a user-friendly interface with flexible features, making it easy for users with a variety of technical backgrounds. Tableau pricing depends on the specific application and solution so, and enables healthcare organizations to choose pricing structures that meet their needs and budget (Tableau, 2023).

Table 3 shows the comparison of **RapidMiner, IBM Watson Health, and Tableau** based on their critical capabilities.

	<b>RapidMiner</b>	<b>IBM Watson Health</b>	<b>Tableau</b>
Key Features	Robust data mining capabilities	AI-powered analytics and insights	Advanced data visualization and analytics
Performance and Scalability	Efficient processing of large datasets	Scalable infrastructure for big data analytics	Efficient processing of healthcare data
Data Access	Support for various data sources and formats	Integration with healthcare data systems	Integration with various data sources
Data Preparation	Preprocessing and cleansing tools for data preparation	Data cleansing and normalization features	Data preprocessing capabilities

Data Exploration and Visualization	Interactive visualization and exploration tools	Rich visualization and dashboarding capabilities	Advanced data visualization and exploration features
Advanced Modelling	Wide range of machine learning algorithms	AI and machine learning models for healthcare analysis	Advanced modeling and predictive analytics
Programming Language	Supports multiple programming languages	Utilizes various programming languages for analysis	Supports multiple programming languages
Operating System	Windows, Mac, and Linux	Cloud-based platform	Windows and Mac
Interfaces	GUI and CLI	Web-based interface and APIs	GUI and web-based interface
Ease of use	User-friendly interface with drag-and-drop functionality	User-friendly interface with guided analytics	User-friendly interface with intuitive features
Price/License	Various pricing options, including free community edition	Pricing based on usage and specific solutions	Pricing based on usage and specific solutions

**Table 3. Comparison of Data Mining Tools in Healthcare**

## CONCLUSION

The study has provided an overview of data mining, data mining techniques, benefits, challenges and limitations. It also discussed the critical issues and challenges associated with data mining tools. It also suggested solutions to help various industries choose the best data mining tool that suited their needs. Finally, and most importantly, several popular data mining tools – SAS Enterprise Miner, KNIME, and R for business, Moodle Learning Analytics, Blackboard Analytics, and Canvas for education, and RapidMiner, IBM Watson Health, and Tableau for healthcare – were described and compared using their critical capabilities to suggest the most suitable tools for the various industries. It is hoped that this approach, i.e. comparing various available tools in domain-specific contexts will be of use to practitioners as well as management in selecting the right data mining tools, and invest in the appropriate training and infrastructure as they plan to leverage data mining in their organizations. Further research can investigate, using a case-study and survey approach, how specific organizations in specific domains undertake the process of tool selection for data mining.

---

## REFERENCES

- Abraham, S. (2019). Data Mining Tools and Techniques: A Review. *International Journal of Computer Science Trends and Technology*, 7(3), 81-85.
- Agaoglu, M. (2016). Predicting instructor performance using data mining techniques in higher education. *IEEE Access*, 4, 2379-2387.
- Baker, R. S. J. D. (2010). Data mining for education. *International encyclopedia of education*, 7(3), 112-118.
- Bhinghe, A. V. (2015). A comparative study on data mining tools (Doctoral dissertation, California State University, Sacramento).
- Blackboard. (2023). *Blackboard*. Blackboard Inc. <https://www.blackboardanalytics.com/>
- Canvas. (2023). *Canvas*. Canvas Network. <https://www.canvas.com/>
- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, 8(6), 866-883.
- Durairaj, M., & Ranjani, V. (2013). Data mining applications in healthcare sector: a study. *International journal of scientific & technology research*, 2(10), 29-35.
- Gordon, J., Perrey, J., & Spillecke, D. (2015). Big data, analytics and the future of marketing and sales. *McKinsey: Digital Advantage*.
- Hill, S. (2020). Evaluating data mining software packages: A review. *Journal of Information Systems Education*, 31(3), 136-144.
- Huang, H., Jeng, J. Y., Chen, Y., & Chou, W. C. (2021). Clinical data mining in healthcare: A systematic literature review. *International Journal of Environmental Research and Public Health*, 18(6), 2914.
- Huang, Y. (2019). The Privacy and Security Risks of Data Mining and Big Data. *Data Science Journal*, 18(1), 27.

Hyperverge. (2022, November 15). *An Overview of Data Mining For Fraud Detection*. Hyperverge. <https://www.hyperverge.co/blog/data-mining-for-fraud-detection>

IBM. (2023). *IBM Watson Health*. Merative.

<https://www.ibm.com/watson-health/merative-divestiture>

Iwatani, E. (2018). Overview of Data Mining`s Potential Benefits and Limitations in

Education Research. *Practical Assessment, Research, and Evaluation*: 23(15).

DOI: Available at: <https://scholarworks.umass.edu/pare/vol23/iss1/15>

Katal, A., Wazid, M., & Goudar, R. H. (2020). Big data: A review of data mining tools and techniques. *Journal of Big Data*, 7(1), 1-34.

Kim, H. (2018). Understanding the Data Mining Techniques: Integration of Different Data Mining Tools. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 2, pp. 1-5).

Knime.(2023). *Knime*. Knime AG. <https://www.knime.com/>

Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of healthcare information management*, 19(2), 65.

Liao, X., & Hu, X. (2020). A survey of big data analytics in healthcare and government. *Journal of Industrial Integration and Management*, 5(3), 2050022.

Linden, A., Kart, L., Hare, J., & Herschel, G. (2016). Critical capabilities for advanced analytics Platforms, *Gartner*.

Linden, A., Krensky, P., Hare, J., Idoine, C., Sicular, S., & Vashisth, S. (2017). Magic Quadrant for Data Science Platforms, *Gartner*.

Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

Mantas, J. (2017). The hazards of data mining in healthcare. *Informatics Empowers Healthcare Transformation*, 238, 80.

Moodle. (2023). *Moodle*. Moodle Pty Ltd. <https://moodlelearninganalysis.com/>

- R. (2023). R. The R Foundation. <https://www.r-project.org/>
- Ramamohan, Y., Vasantharao, K., Chakravarti, C. K., & Ratnam, A. S. K. (2012). A study of data mining tools in knowledge discovery process. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(3), 191-194.
- Rana, S. (2022). Understanding the Importance of Data Mining Skills for Business Growth. *International Journal of Business Data Communications and Networking (IJBDCN)*, 18(2), 1-12.
- RapidMiner. (2023). *RapidMiner*. An Altair Company. <https://rapidminer.com/>
- Rawat, K. S., & Author, C. (2017). Comparative Analysis of Data Mining Techniques, Tools and Maching Learning Algorithms for Efficient Data Analytics. *JOSR J. Comput. Eng*, 19(4), 56-60.
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, 3(1), 12-27.
- Saabith, A. L. S., & Fareez, M. M. M. (2018). Comparative Analysis of Various Tools for Data Mining and Big Data Mining. *International Journal Of Engineering Research & Technology*, 7(11).
- Santos-Pereira, J., Gruenwald, L., & Bernardino, J. (2022). Top data mining tools for the healthcare industry. *Journal of King Saud University-Computer and Information Sciences*, 34(8), 4968-4982.
- SAS Enterprise Miner. (2023). *SAS Enterprise Miner*. SAS Institute Inc. [https://www.sas.com/en\\_us/software/enterprise-miner.html](https://www.sas.com/en_us/software/enterprise-miner.html)
- Schultz, K. (2021). An evaluation of data mining software tools. *Journal of Global Research in Computer Science*, 12(1), 20-23.
- Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, 42(1), 85-106.
- Şimşek, H. (2023, April 10). *Data Mining: What is it & Why do businesses need it in 2023?* AIMultiple. <https://research.aimultiple.com/data-mining/#what-are-the-disadvantages-of-data-mining>

- Son, L. H., & Fujita, H. (2019). Neural-fuzzy with representative sets for prediction of student performance. *Applied Intelligence*, 49(1), 172-187.
- Stahl, F., Gabrys, B., & Weiss, G. (2022). Explainable and interpretable data mining and knowledge discovery: State of the art and emerging trends. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(2), e1488.
- Tableau. (2023). *Tableau*. Tableau Software, LLC, A Salesforce Company. <https://www.tableau.com/>
- Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to data mining. *Pearson Education India*.
- Ünal, F. (2020). Data mining for student performance prediction in education. *Data Mining-Methods, Applications and Systems*, 28, 423-432.
- Verma, S. (2020). Selection of Data Mining Tool: An Overview. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(6), 4146-4150.
- Yang, F., Lee, J. H., & Wang, F. (2020). Artificial intelligence and big data analytics in cyber security. *Journal of Big Data*, 7(1), 1-34.
- Zhang, M., Fan, J., Sharma, A., & Kukkar, A. (2022). Data mining applications in university information management system development. *Journal of Intelligent Systems*, 31(1), 207-220.