

Data Mining and Machine Learning: A Comparative Survey in Choosing the Right Approach

Dr D Suresh Babu¹ and Dr K Aruna²

*¹Dr D Suresh Babu, Head Department of Computer Science & Applications
SR & BGNR Government Arts & Science College, Khammam, Telangana India*

*²Dr K Aruna Lecturer, Department of Computer Science
CKM Arts & Science College, Warangal, Telangana, India*

Abstract

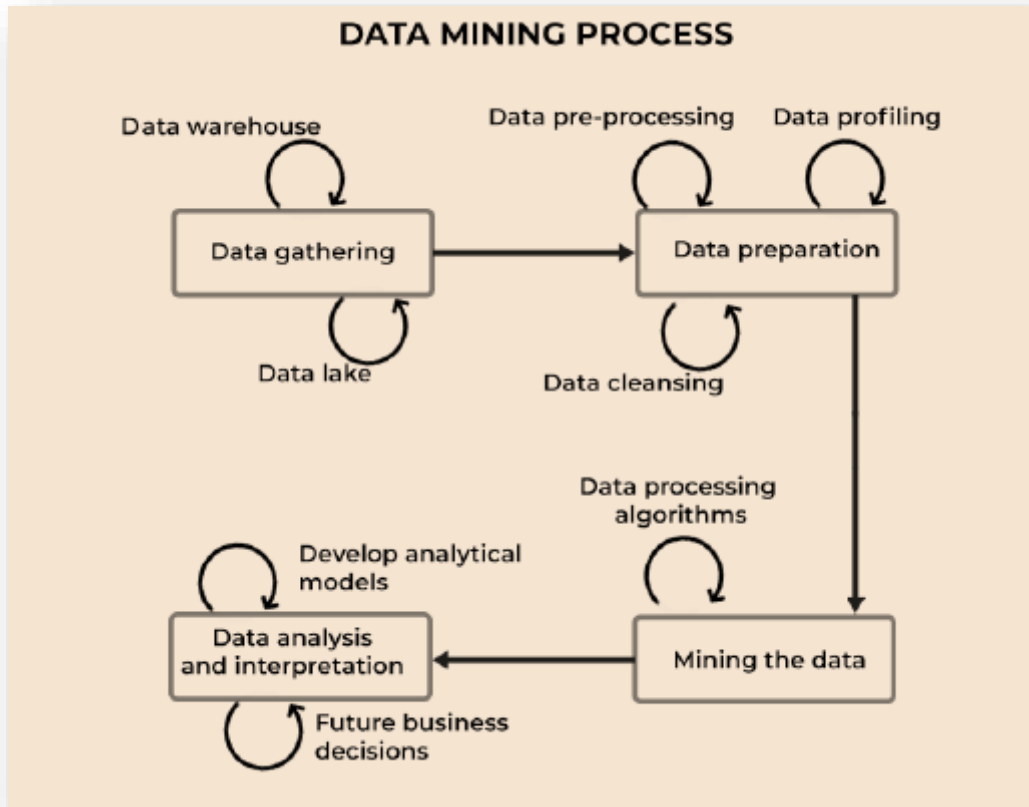
Data Mining and Machine Learning are two strongly interconnected however divergent fields in data analysis. With both techniques extracting important insights, it becomes decisive to realize their characteristics, applications, and methodologies. This paper aims to throw the light on aspects of Data Mining and Machine Learning and how do they differ in terms of goals and approaches. These questions, quickly exploring the key differences and overlaps between data mining and machine learning. By separating their distinctions, we can better take hold of their potential and make informed decisions using these dominant analytical tools.

Keywords: Algorithm, Block Chain, Career, Data Analysis

Data Mining:

Data mining, sometimes called the discovery of knowledge in databases, analyzes enormous amounts of data from multiple data sets to collect significant knowledge that helps businesses resolve problems, predict patterns, reduce pitfalls, and uncover new opportunities. Data miners filter through plenty of data in looking for useful components and materials, similar to what miners do in real mining operations.

Defining an organization's goal is the first step in the data mining approach. subsequent that, information is gathered from various sources and added to databases, which act as reservoirs for data analysis. Data cleaning entails filling any gaps in data and eliminating duplicates, and finding data patterns using sophisticated methods and mathematical frameworks.



Machine Learning:

Machine Learning is a way that seeks to make computers more like human beings in their behavior and judgments by allowing them to achieve knowledge and write their code. The Machine Learning approach is automated and sophisticated, based on the experiences of the machines throughout the process.

Machine learning is a data mining method that focuses on developing algorithms to increase the usability of data-derived experiences. It is a function of a system to gain insight from a targeted data set, whereas data mining uses methods created by machine learning to predict outcomes.

Variation Between Data Mining and Machine Learning Techniques:-

There are three types of machine learning:

Supervised Machine Learning:-

This meticulous type of machine learning integrates past inputs. It results in machine learning algorithms interpreting every input/output combination that enables the

algorithm to adjust the predictive model to produce outcomes as closely equivalent to the estimated outcome as feasible. Neural networks, decision trees, linear regression, and support vector machines are basic supervised learning techniques.

Unsupervised Machine Learning:-

This type of machine learning is very much advantageous when you require it to find trends and use the data for making conclusions. Hidden Markov models, k-means, hierarchical clustering, and Gaussian mixture models are common unsupervised learning algorithms.

Reinforcement Machine Learning:-

Reinforcement learning teaches a computer to respond suitably and maximize its benefits in certain circumstances. It generates actions and rewards using a mechanism and a setting, and the process has a beginning and an ending. Deep adversarial networks, Q-learning, and temporal differences are common algorithms.

Techniques Used in Data Mining:-

The techniques majorly used in data mining are as follows:

1. **Classification:** By implementing this technique, one can gather essential and relevant data and metadata details. This data mining procedure facilitates categorizing data into several groups.
2. **Clustering:** Data mining techniques like clustering analysis finds comparable data. This method enables the identification of the variations and commonalities among the data.
3. **Regression:** Regression analysis is the data mining technique applied to discover and evaluate relationships among elements because of adding the other component.
4. **Outer:** This kind of data mining technique refers to discovering data points in the data set which differ from a typical trend or predicted behavior.
5. **Sequential Pattern:** The sequential pattern is a method of data mining used for detecting recurring trends by examining sequential data. Finding intriguing segments among a group of sequences is what it entails. The significance of a sequence is often determined by its length, frequent occurrence, and other factors.
6. **Prediction:** Prediction utilizes several data mining techniques, including trends, clustering, classification, etc. To forecast a future event, it appropriately sequences the analysis of past events or instances.
7. **Association Rules:** Association rules are if-then statements that can help to illustrate the likelihood of interactions among data elements inside vast collections of information in many different kinds of databases.

Data Mining Tools

The most popular tools used in data mining are as follows:

- Orange Data Mining
- SAS Data Mining
- Datamelt Data Mining
- Rattle
- Rapid Miner
- Oracle Data mining
- IBM SPSS Modeler
- Weka
- Apache Mahout
- TeraData

Data Mining and Machine Learning Applications:**Applications of Data Mining:-**

Some of the applications of data mining are as follows:

1. For enhancing healthcare systems, data mining offers a lot of potential. It highlights best practices for utilizing insights and data to improve care and reduce expenses.
2. Data mining tools in banking could be the ideal solution due to their ability to discover trends, damage, market challenges, and other interactions that managers must be aware of.
3. The “educational data mining” field is expanding swiftly and involves developing methods for extracting information from data collected in educational settings.
4. The methods used for conventional fraud detection are laborious and challenging. Data mining helps in the conversion of data into insights and the discovery of important patterns.
5. Data mining enables organizations to divide their customer base into distinct segments and customize services to meet each group’s unique needs.

Applications of Machine Learning:

Some of the applications of machine learning are as follows:

1. One of the most popular uses of machine learning is image identification. It identifies things like digital photos, people, places, and items.
2. Amazon, Netflix, and other e-commerce and entertainment businesses commonly utilize machine learning for recommending products to users.
3. Machine learning makes our online transactions safe and secure by identifying fraudulent transactions.
4. Machine learning identifies diseases. As a result, medical technology is developing rapidly and can now create 3D models capable of determining the exact spot of lesions within the brain.

5. Sentiment analysis uses an instantaneous form of machine learning to predict the sentiment or viewpoint of the speaker or writer.

Advantages and Disadvantages of Data Mining and Machine Learning

Advantages of Data Mining

1. Governments, businesses, and organizations can acquire reliable details through data mining.
2. Data mining finds fraud and challenges that standard data analysis techniques might miss.
3. Finding variations and patterns in user activity can be done through data mining.

Disadvantages of Data Mining

1. Data mining occasionally fails to produce reliable information.
2. Large databases are necessary for effective data mining.
3. Data mining is often an extremely costly operation.

Advantages of Machine Learning

1. Machine learning can review large quantities of data, identifying certain patterns and trends that individuals might miss.
2. Machine learning algorithms are adept at managing multidimensional and multivariate data in variable or unpredictable contexts.
3. Specific procedures can be automated by machine learning algorithms, which lower labor costs and free organizations from concentrating on other value-adding activities.

Disadvantages of Machine Learning

1. Machine learning algorithms are resource-intensive and computationally demanding.
2. It requires time and effort to train a machine-learning algorithm.
3. ML is self-sustaining but vulnerable to errors.

Key Differences between Data Mining and Machine Learning

Some of the differences are to be considered are:

Parameters	Data Mining	Machine Learning
Definition	It is the technique of discovering significant patterns from huge datasets	It is the method of organizing and interpreting unstructured data to produce meaningful data and direction

Purpose	The major purpose of data mining is to enhance the usability of the data used presently	Data analysis is carried out to generate hypotheses, which ultimately results in the generation of pertinent data to support company decisions
Techniques and tools used	Data mining is more of a research activity that employs techniques such as machine learning. Tools used: Rattle, Rapid Miner, Oracle Data Mining, etc	It is an independent and trained system that does the work precisely. Tools used: Excel, Power BI, Tableau, etc.
Data types used	Transactional data, Data warehouse and data stored in databases	Nominal, Ordinal, Discrete and Continuous.
Applications	It is employed in cluster analysis, and the information is extracted from the data warehouse.	It reads machinery and is applied to computer design, spam filtering, fraud detection, and web search

Similarities between Data Mining and Machine Learning:

Some of the similarities between them are as follows:

- Machine learning and data mining have both been implemented in predictive modeling. Sentiment analysis is a related application
- They include statistics, mathematical concepts, and algorithms
- They also filter across data, various tools, and applications using algorithmic methods
- They sometimes adopt comparable structural or algorithmic methods

Use Cases of Data Mining and Machine Learning

Data mining techniques extract new insights from existing data or anticipate the outcome using past data. Data mining's limitations are solved by machine learning, which enables it to develop much more efficiently. Additionally, machine learning can address problems independently because it is more precise and not as prone to errors.

However, it is vital to keep up with the data mining process because it will help to identify the challenge of a certain organizational structure. For businesses to succeed and collaborate more effectively, data mining and machine learning are essential.

Some of the use cases which can establish **data mining and machine learning** are as follows:

Data Mining

- **Data Mining in Finance:** Facilitates locating hidden connections between various financial metrics that are required to find unusual activity with an elevated risk. It generally distinguishes between fraudulent and corrective behavior as it

has been done by gathering historical facts and then transforming them into beneficial factual information.

- **Data Mining in Crime and Intelligence:** Enhances detection of anomalies while improving detection of intrusions and spotting dubious behavior promptly. Text-based crime reports would be converted into document types. That could help the process of matching crimes.
- **Data Mining in Marketing:** Predicting a customer's behavior to guide customized loyalty programmes is feasible by studying the links between criteria like age, gender, and preferences. Data mining in marketing can also forecast which consumers are most likely to discontinue a service, what attracts them depending on their searches, and what information should be included in a mailing list to increase response rates.

Machine Learning

- **Machine Learning in Stock Market:** Organizations worldwide use machine learning methods and models for forecasting stock market prices by analyzing sentiment. Social media is one of the many data sources where you can perform sentiment analysis. The use of classification and clustering techniques, together with NLP, enables the classification of stocks into three groups: negative, positive, or neutral.
- **Machine Learning in Dynamic Pricing:** Dynamic pricing can be accomplished with the help of machine learning algorithms, which can also be significant in increasing profits and returns. ML techniques under supervision can pick up new patterns based on the provided data. The outcomes of such algorithms may be updated regularly to comply with trends. Online stores estimate the dynamic pricing of a good or service using ML algorithms and methodologies.
- **Machine Learning in Image Recognition:** Using machine learning, it is possible to train applications to recognize objects and other photo elements. A comprehensive library of photographs is analyzed pixel by pixel using a neural network. After verifying their own information, each neuron provides insight, and the network combines millions of these findings into a coherent analysis. Using an open image database, developers train machine learning algorithms to recognize these photos.

Conclusion: Which One to Choose:

Data mining and machine learning are complementary yet distinct disciplines that help businesses extract meaningful data. While data mining focuses on uncovering hidden patterns and relationships within data, machine learning goes beyond building predictive models and making automated decisions. Understanding the nuances between these approaches is essential for effectively applying them in real-world scenarios.

To explore deeper into the details of data mining and machine learning, consider enrolling in our BlackBelt Program. This broad program offers in-depth training, hands-

on experience, and practical knowledge to enhance your skills in data analysis, predictive modeling, and advanced machine learning techniques. Take the next step towards becoming a proficient data scientist and leverage the power of data mining and machine learning to drive meaningful insights and impactful decisions.

References:

- [1] Data Mining in the Factual World: Issues, Challenges and Recommendations: Advances in Computational Sciences and Technology, Volume 4 Number 3 (2011) pp. 279-290.
- [2] Heart disease prediction using machine learning techniques: Open Access Research Journal of Engineering and Technology, 2022, 03(01), 001–006.
- [3] Web Usage Mining: A Research Concept of Web Mining: International Journal of Computer Science and Information Technologies, Vol. 2 (5), 2011, 2390-2393
- [4] *Analysis on Periodic Web Personalization for the Efficiency of Web services: International Conference on Inventive Communication and Computational Technologies (ICICCT): IEEE Explore, DOI: 10.1109/ICICCT42250.2018*
- [5] Machine-Learning and Deep-Learning Techniques for Air Quality Analysis: Journal of Information and Computational Science, Volume 11 Issue 12 – 2021.
- [6] Prediction of Heart Disease with Advanced Machine Learning Techniques: International Journal of Advances in Engineering and Management (IJAEM) Volume 4, Issue 7 July 2022, pp: 709-713.
- [7] *Machine Learning: Algorithms, Real-World Applications and Research Directions: <https://doi.org/10.1007/s42979-021-00592-x>*