# Data Mining and Its Applications for Knowledge Management: A Literature Review

**Ramesh Lavhe[#1], Sayali Kokane[*2], Ashok Kalal[#3], R. A. Nikam[#4]**
rajshri.nikam@abmspcoerpune.org

***Article Info***
***Page Number: 244 - 248***
***Publication Issue:***
***Vol 68 No. 1 (2019)***

***Article History***
***Article Received:*** *09 September 2019*
***Revised:*** *16 October 2019*
***Accepted:*** *21 November 2019*
***Publication:*** *28 December 2019*

**Abstract**

Data mining is a prominent area of knowledge management and is one of the most critical phases in the process of knowledge discovery in databases. It is also one of the most crucial steps in the process. In the future decades, research in data mining will continue to expand in both the commercial world and the world of learning organizations. The purpose of this review article is to investigate the uses of data mining methods that have been developed to help the process of knowledge management. After that, the logic behind knowledge management and the primary knowledge management instruments that are included into the knowledge management cycle are outlined. In conclusion, a discussion and summary of the many uses of data mining methods throughout the process of knowledge management are presented.

**Keywords**: -Data mining; Data mining applications; Knowledge management

## Introduction

Data mining is a strong approach that has recently come into its own as a method for gleaning useful insights and information from massive amounts of data. Within the framework of knowledge management, data mining provides a variety of applications that, when used appropriately, may increase organizational efficiency, encourage innovation, and enhance decision-making. This literature review presents a complete study of the research that has already been conducted on data mining and the applications it has for knowledge management. It investigates a wide variety of data mining approaches, algorithmic frameworks, and methodological approaches that are used in knowledge management domains. In addition, the paper investigates the most important advantages, problems, and potential future directions associated with using data mining in order to implement efficient knowledge management systems.

## 2.0 Literature Survey

### A. Overview of data mining and knowledge management.

The process of identifying patterns, correlations, and insights hidden within enormous amounts of data is referred to as data mining. It entails making use of a number of different statistical and machine learning approaches in order to extract useful knowledge and information. Methods for data mining include, among others, classification, clustering, association rule mining, and anomaly detection. Data mining allows companies to make educated choices, see patterns and trends, and obtain insights that can be put into action since it involves the analysis and interpretation of data. Management of Knowledge: Management of knowledge is the methodical collection, organization, and application of an organization's accumulated expertise and experience. It places an emphasis on the creation of an atmosphere conducive to the exchange of information, the archiving of that information, and the use of that information to improve decision-making, stimulate innovative thinking, and improve organizational performance. The procedures, techniques,

and technologies that are used to acquire, preserve, and distribute knowledge assets are referred to collectively as "knowledge management." Its ultimate goal is to improve overall productivity and competitiveness by easing the process of creating new information, sharing it across people, teams, and departments, and applying it to new situations.

The relationship between Data Mining and Knowledge Management Data mining is an integral part of knowledge management because it enables businesses to get valuable insights and information from their data. This is an essential component of knowledge management. Organizations are able to unearth previously hidden patterns, trends, and linkages within their data repositories if they employ methods from the field of data mining. These new perspectives have the potential to be developed into important knowledge and integrated into systems for the administration of that knowledge. The mining of data improves the effectiveness of information discovery, lends support to the processes of decision making, and contributes to the development of new solutions. It gives the capability to find important information that can be shared and exploited throughout the firm, promoting a culture of continual learning and progress in the process..

**B. Importance of data mining in knowledge discovery and decision-making**

The mining of data is an important step in both the discovery of new knowledge and the process of making decisions since it provides both useful insights and information that can be put to use. The following is a list of important factors that illustrate the significance of data mining in different areas:

- **Pattern and Relationship Identification:**The mining of data is an important step in both the discovery of new knowledge and the process of making decisions since it provides both useful insights and information that can be put to use. The following is a list of important factors that illustrate the significance of data mining in different areas:
- **Predictive Analytics:** Mining for data opens the door to predictive modelling, which in turn enables businesses to make educated choices based on what they anticipate will happen in the future. Data mining algorithms may develop predictive models that estimate future events or behaviours by analyzing previous data and recognizing patterns. This process is known as pattern recognition. Because of this, businesses are better able to foresee possible dangers, opportunities, and trends in the market, which enables them to make proactive decisions.
- **Improved Resource Allocation:** The mining of data gives businesses the ability to maximize the effectiveness of their resource allocation and use. It is possible for businesses to successfully allocate resources by making choices that are driven by data after doing an analysis of data relating to consumer preferences, market trends, and operational procedures. This comprises the optimization of workforce numbers, administration of supply chains, control of inventories, and allocation of budgets. The insights that may be discovered via data mining provide firms the ability to more effectively manage resources and generate better results from their operations.
- **Personalization and Customer Segmentation:** Data mining gives businesses the ability to better understand their consumers and to customise the goods and services they provide as a result. Companies are able to segment their client base and customize their services to certain consumer groups by doing an analysis of customer data, which may include the customer's purchasing history, preferences, and behavior. This ultimately results in higher levels of consumer satisfaction, more effective targeted marketing initiatives, and increased customer loyalty.
- **Fraud Detection and Risk Management:** Methods of data mining are quite helpful in the process of identifying fraudulent activity and controlling risks. Data mining algorithms are able to uncover suspicious patterns or anomalies by analysing enormous amounts of data. These patterns or anomalies may signal fraudulent behaviour or possible hazards. Because of this, firms are better able to take preventative actions against fraud, reduce risks, and safeguard their assets.
- **Strategic Decision-Making:** The mining of data may provide insightful knowledge that can help in the process of making strategic decisions. Organizations are able to make educated judgments on product development, market growth, pricing strategies, and competitive positioning if they conduct in-depth analyses of market trends, data on their competitors, and internal performance indicators. These insights, which are driven by data, guarantee that choices are founded on facts and are in line with the objectives of the business.

## 3.0 System Study

The research objectives and scope of the data mining study can be defined as follows:

**Research Objectives:**

- To investigate and analyse a variety of data mining methods and algorithms.
- To determine the many uses of data mining and the advantages it provides in a variety of fields.
- To conduct an investigation of the difficulties and constraints that are involved with data mining.
- To explore the influence that data mining has on the discovery of new information and the process of making decisions.
- To investigate the ramifications of data mining on privacy and the ethical concerns that accompanies it.

**Scope of the Study:**

- Classification, clustering, association rule mining, and anomaly detection are just some of the topics that will be covered in this study on data mining methods.
- It investigates how data mining may be used in fields such as healthcare, finance, marketing, and customer relationship management.
- It covers an examination of the obstacles and constraints in data mining, such as data quality, scalability, and interpretability.
- The research analyses both structured and unstructured data sources for the analysis of data mining.
- The purpose of this research is to investigate the effects that data mining has on the discovery of new information, the effectiveness of decision support systems, organizational efficiency, and creativity.
- Ethical aspects and privacy implications associated with data mining are highlighted, including the ethical use of data and maintaining compliance with privacy legislation

## 4.0 System Design

When it comes to gleaning useful insights and information from enormous datasets, the applications of data mining methods and algorithms are very essential. The following are some data mining approaches and algorithms that are regularly used:

**A. Classification:**

- **Decision Trees:** A tree-like model where each internal node represents a feature, each branch represents a decision rule, and each leaf node represents a class label.
- **Naive Bayes:** Based on Bayes' theorem, this algorithm assumes that features are conditionally independent given the class label.
- **Support Vector Machines (SVM):** A supervised learning algorithm that finds a hyperplane to separate different classes.

**B. Clustering:**

- **K-Means:** Divides data into non-overlapping clusters by minimizing the sum of squared distances between data points and the centroid of each cluster.
- **Hierarchical Clustering:** Builds a hierarchy of clusters by merging or splitting them based on distance or similarity measures.
- **DBSCAN**: Density-based spatial clustering algorithm that groups together data points based on density connectivity.

**C. Association Rule Mining**:

- Apriori: Discovers frequent itemsets by generating candidate itemsets and pruning based on minimum support threshold.
- FP-Growth: Constructs a compact data structure (FP-tree) to mine frequent itemsets without generating candidate itemsets.

**D. Regression:**

- Linear Regression: Predicts a continuous numerical value by fitting a linear relationship between the dependent variable and independent variables.
- Decision Trees for Regression: Similar to decision trees for classification, but the leaf nodes contain continuous values instead of class labels.
- Support Vector Regression (SVR): Applies SVM principles to regression problems, finding a hyperplane that has a maximum number of data points within a certain margin.

**E. Anomaly Detection:**

- **Density-Based Anomaly Detection:** Identifies anomalies as data points that deviate from the expected density distribution.
- **Isolation Forest:** Constructs random decision trees to isolate anomalies that require fewer partitions.
- **One-Class Support Vector Machines (OC-SVM):** Learns the boundary of normal instances and classifies instances outside the boundary as anomalies.

**F. Neural Networks:**

- **Multilayer Perceptron (MLP):** Consists of multiple layers of interconnected neurons for learning complex patterns.
- **Convolutional Neural Networks (CNN):** Designed for analyzing visual data, using convolutional layers for feature extraction.
- **Recurrent Neural Networks (RNN):** Suitable for sequential data, with connections between nodes forming a directed graph.

## Acknowledgement

We would like to express our sincere gratitude to Er. AtulMarathe (Adjunct Professor), Er. Ashok Saraf (Innovation Club Member) for their invaluable guidance and support throughout the research process. We also wish to thank Dr. Sunil B. Thakare (Principal, APCOER, Pune) for their support. Finally, we are grateful to all of the research participants who generously gave their time and effort to this project.

### References

[1] Han, J., &Kamber, M. (2006). Data mining: concepts and techniques. Morgan Kaufmann.
[2] Witten, I. H., Frank, E., & Hall, M. A. (2016). Data mining: practical machine learning tools and techniques. Mor Dhabalia, D. (2019). A Brief Study of Windopower Renewable Energy Sources its Importance, Reviews, Benefits and Drwabacks. Journal of Innovative Research and Practice, 1(1), 01–05.
[3] Mr. Dharmesh Dhabliya, M. A. P. (2019). Threats, Solution and Benefits of Secure Shell. International Journal of Control and Automation, 12(6s), 30–35.
[4] Verma, M. K., & Dhabliya, M. D. (2015). Design of Hand Motion Assist Robot for Rehabilitation Physiotherapy. International Journal of New Practices in Management and Engineering, 4(04), 07–11.gan Kaufmann.
[5] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37-54.
[6] Aggarwal, C. C., &Zhai, C. (2012). Mining text data. Springer Science & Business Media.

[7] Chen, M., Han, J., & Yu, P. S. (2016). Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6), 866-883.
[8] Mitchell, T. M. (1997). Machine learning. McGraw-Hill.
[9] Tan, P. N., Steinbach, M., & Kumar, V. (2005). Introduction to data mining. Pearson Addison Wesley.
[10] Larose, D. T. (2014). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons.
[11] Chawla, N. V., Bowyer, K. W., Hall, L. O., &Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357.
[12] Zhang, Z., &Oommen, B. J. (2007). A survey on data clustering. Pattern Recognition, 40(8), 2570-2600.