Application of Data Mining in Knowledge Management: A Review

Jovanne C. Alejandrino
Institute of Computing, Davao del Norte State College, New Visayas, Panabo City, Philippines
jovanne.alejandrino@dnsc.edu.ph

ABSTRACT

Data mining is one of the essential tools to capture knowledge. This paper explores the applications of data mining techniques that have been used in the knowledge management process. The journal articles that are published in different international journal publications are analyzed and classified. This paper first briefly describes the concept of data mining and knowledge management. The discussion on the result is divided into four topics: (1) knowledge resource, (2) knowledge datasets, (3) data mining tasks, and (4) data mining algorithms. Lastly, the applications of data mining in the knowledge management process are summarized and discussed.

Key words: knowledge management, data mining, knowledge discovery, data analysis, application of data mining, data mining algorithms, knowledge management process

1. INTRODUCTION

In recent days, knowledge is referred to as a key factor in the nation's progress. Knowledge-based urban development refers to the development of a city driven by a knowledge economy that fuels economic growth [1]. Knowledge can be generated from the information, and information can be produced through the processing of data [2]. Many institutions have collected and stored a massive amount of data. However, the transformation of data into valuable and useful information is needed to discover helpful knowledge from the data collected. Managing knowledge from data is challenging. Information technology helps employ knowledge management in organizations to assist, share, integrate, and distribute knowledge. Knowledge Discovery from data (KDD) or referred to as data mining, is the process of discovering and extracting relevant information out from the amount of data collected [3].

Data Mining, as defined by Gartner Group, is "the process of discovering significant new patterns, correlations, and trends by filtering through large amounts of data stored in repositories and by the application of pattern recognition technologies as well as statistical and mathematical techniques" [4]. In its simplest form, data mining doesn't involve looking for specific information. Rather than starting from a question or a hypothesis, data mining finds patterns that are already present in the data.

Every enterprise can benefit from collecting and analyzing its data. The medical industry uses its data to detect trends and anomalies in its patient records. Search engines use their data to do better ranking and advertisement placements. The environmental and public health sectors can spot patterns and anomalies in their data to improve decision-making. Data mining can be used in cybersecurity and computer network intrusion detection, pattern analysis in bioinformatics and pharmaceutical data, monitoring and conserving of energy, financial and business intelligence data, spotting trends in social networking sites, and many more [3].

The problem of several industries is how to analyze their data. As part of knowledge management research, this paper focuses on surveying data mining applications in knowledge management through a literature review. The purpose of this paper is to identify how data mining has emerged and is used in knowledge management and how it plays a vital role in the organization's progress.

For article filtering, this paper used keywords like "knowledge management," "data mining," and "knowledge discovery" in the article title, abstract, and author keywords fields in Science Direct and Google Scholar databases. The full text of the article is carefully read and reviewed to remove those articles that are not related to the application of data mining in Knowledge Management and not mentioned how data mining could be employed in Knowledge Management. There are only ten articles selected for the selection criteria.

Based on the ten articles on data mining applications in knowledge management, this paper classifies according to the six categories of data mining methods, which are classification, clustering, neural network, association estimation, sequence-based analysis, and visualization.

The purpose of this review paper is to review related literature on the application of data mining techniques for Knowledge Management. The sequence of this paper as follows: first, data mining definition; second, the definitions of knowledge management are presented; third, articles about data mining in Knowledge Management are analyzed, and the findings of the classification are reported; and lastly, the conclusions of the study are discussed.

2. METHODOLOGY

The databases Google Scholar, Science Direct, IEEE, ResearchGate,as shown in Figure 1, were searched from Jan 2010-Dec 2020. The search terms used were as follows: "data mining," "knowledge management," "knowledge discovery," "Big Data," "Data Science." These terms were used using the Boolean operators "and" and "or." Searches were limited to English language studies only. Selected studies provided primary reference lists were checked, and hand searches of relevant journals were conducted. Eighteen articles were reviewed independently, and the final selection (n = 10) was agreed on following debate and discussion..
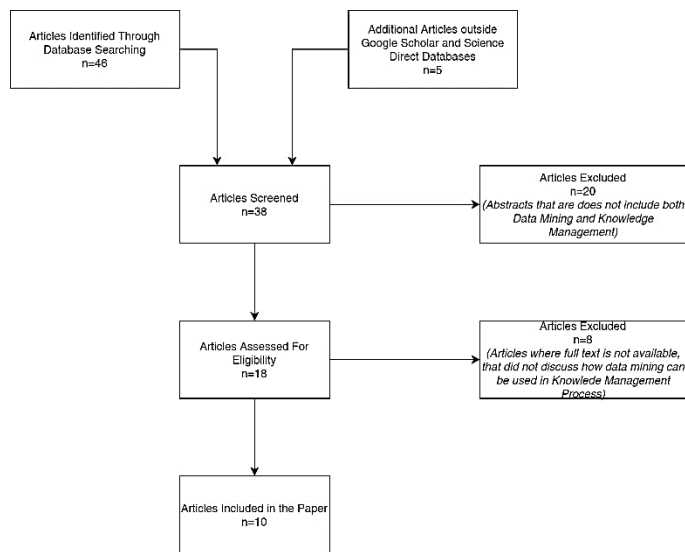

Figure 1. Article Filtering Process

3. DATA MINING

3.1 Definition of Data Mining

Many terms have been associated with data mining, like knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. But among these terms, Knowledge Discovery from data (KDD) is the most popular term [3]. KDD, often called data mining, purposely aims to find useful information from extensive collections of data. The main functions of data mining are the application of various techniques and algorithms to discover and extract patterns of stored data. The sources of these data include databases, data warehouses, the Web, other information repositories, or data that run into the system dynamically [3][4]. Figure 2 shows the process of data mining consists of an iterative sequence method [3].
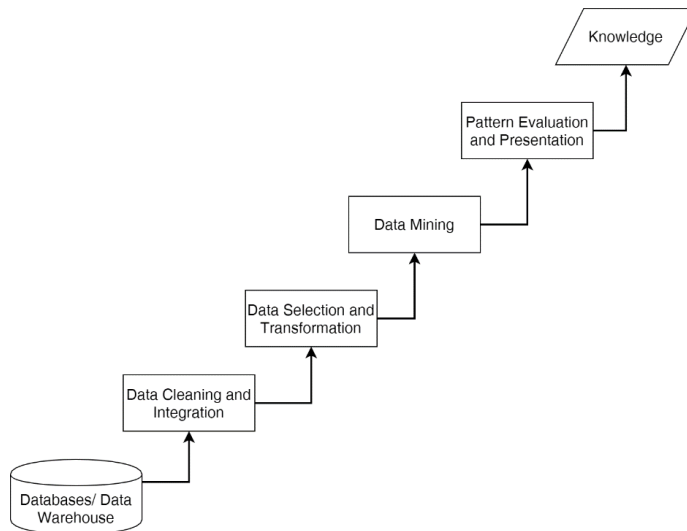

Figure 2. Data Mining Process

Based on Figure 2, data mining is consisting of an iterative sequence technique as follows:

1. *Data Cleaning:* Removing noise and inconsistent data, combining multiple data sources.
2. *Data integration:* Combining multiple data sources.
3. *Data Selection*: Selecting relevant data that are retrieved from the database needed in the analysis.
4. *Data Transformation*: Transforming data into appropriate forms to perform data mining.
5. *Data mining*: Choosing of data mining algorithm thatapplies to the pattern in the data.
6. *Pattern evaluation:* Identifying the interesting patterns representing knowledge.
7. *Knowledge presentation:* Translating useful patterns into terms that humanscan understand.

3.2 Data Mining Tasks

[5]define six primary functions of data mining:

1. *Classification*: The process of finding a model that describes and distinguishes data classes or concepts.
2. *Regression*: The process of mapping a data item to a real-valued prediction variable.
3. *Clustering*: The process of identifying a finite set of categories or clusters to describe the data.
4. *Dependency Modeling (Association Rule Learning).* The process of discovering a model that describes significant dependencies between variables.
5. *Deviation Detection (Anomaly Detection):* The process of discovering the most significant changes in the data.
6. *Summarization:* The process is finding a short description of a subset of data.

There are two primary objectives of data mining- prediction and description. Prediction involves using some variables in data sets to foresee unknown values of other significant variables (e.g., classification, regression, and anomaly detection). Description involves discovering human-understandable patterns and trends in the data (e.g., clustering, association rule learning, and summarization) [6].

4. KNOWLEDGE MANAGEMENT

4.1 Definition of Data Mining

Several definitions of knowledge management (KM) can be found in different pieces of literature written by many research scholars. It is essential first to understand the meaning of knowledge before having a better understanding of knowledge management.

Knowledge is a comprehensive concept with deep meanings, bearing the belief that it increases an organization's ability for effective action. Knowledge can be further divided into two kinds, that is, tacit knowledge and explicit knowledge [7].

According to [8], knowledge management is the formalization of and access to experience, knowledge, and expertise that create new capabilities, enable superior performance, encourage innovation, and enhance customer value. The authors also describe knowledge management as an umbrella term for a variety of interlocking terms, such as knowledge creation, knowledge valuation and metrics, knowledge mapping and indexing, knowledge transport, storage and distribution, and knowledge sharing.

The knowledge management process as shown in the Figure 3 focuses on knowledge acquisition and application, knowledge capture and creation, knowledge and knowledge sharing, and dissemination (Figure 2) [9]. The process of acquisition, creation, sharing, and distribution, and application of knowledge can be facilitated by information technology.
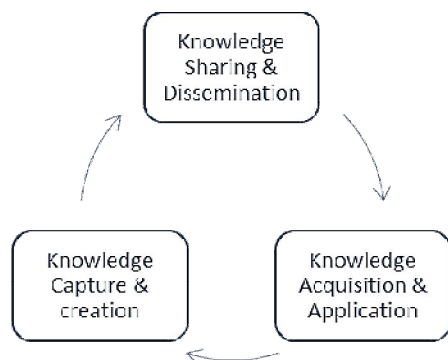

Figure 3. Knowledge Management Cycle

4.2 Knowledge Management Tool

This section shows an overview of different KM technologies as a tool to capture and create knowledge. In the study of [10] about mining customer knowledge for the product line and brand extension in retailing, it used seven categories to classify KM:

1. KM Framework.
2. Knowledge-based systems
3. Data Mining
4. Information and Communication Technology
5. Artificial Intelligence (AI) / Expert Systems (ES)
6. Database Technology
7. Modeling

Data mining is an instrument for knowledge discovery. Data mining is one of the tools used in the Knowledge management process. Data mining can be helpful for KM in the two following main aspects [11]:

1. To share a common understanding of the context of Business Intelligence among data miners.
2. To use Data Mining as a substantial tool to extend human knowledge.

5. KNOWLEDGE MANAGEMENT

The review of the selected articles discussed the applications of data mining to organizational knowledge management for capturing, storing and retrieving, and sharing knowledge. The reviewed articles are divided into four main groups: (1) knowledge resource, (2) Datasets, (3) data mining tasks, and (4) data mining techniques and applications used in KM. Table 1 shows how the articles are categorized and distributed.

5.1 Knowledge Resources

This section identifies different sectors that utilize data mining in the knowledge management process.

1. Healthcare/Medical: Health and medical organizations used DM to utilize knowledge in predicting health insurance fraud and under diagnosed patients, and identify and classify at-risk people in terms of health intending to reduce healthcare cost; DM introduced how data mining technologies have been used for several purposes, including research and study in the biomedical and healthcare fields [12].
2. Energy Conservation: Businesses use data mining technology to generate knowledge in the analysis of building energy consumption data, including energy consumption prediction, fault diagnosis, optimal operation [13].
3. Education: A new area of research that uses techniques of data mining is known as Educational Data Mining. The education sector uses machine learning algorithms and statistical methods to sight knowledge to help for the interpretation of student's

learning habits, academic performances, and further improvements- if necessary [14].

4. Banking/Finance: Leading banks are using Data Mining (DM) tools to gain knowledge for the customer and product segmentation and profitability, credit approval, credit scoring, predicting payment default, marketing, fraud detections in real-time, targeting and acquiring new customers, and analysis of the customers" purchase patterns over time for better customer retention and relationship [15].

5. Retailing: Knowledge extraction by data mining in the retailing industry sights knowledge by clustering the segmentation for the possible product line and brand extension to identify the market to customer clusters [10].

6. Cybersecurity: Data mining is being used to cybersecurity in addressing security problems such as intrusion detection and auditing. Anomaly detection methods are used to detect unusual patterns and behaviors. Analyzing audit data, conduct an analysis using various data mining tools to see if there are potential anomalies in accessing the data repository or a data warehouse [16].

7. Collaboration and Team Work: Many organizations used data mining techniques to improve collaboration and teamwork processes. Data mining tools are used to extract knowledge from collaboration patterns from system logs and analyze the impact of collaboration patterns on process efficiency. Data mining identifies groups of workers with similar task-related information needs that are based on the similarities of workers' knowledge flow [17].

8. Agriculture: Data Mining has been widely utilized by the agricultural sector to sight reliable and timely forecasts of crop production for several policy decisions relating to storage, distribution, pricing, marketing, import-export, etc. [18].

9. Customer Relationship Management (CRM): Companies utilize data mining solutions to prevent both losses of existing customers and attracting new customers. Data mining is used in the Customer Profiling System (CPS), resulting in the early detection of customers who tend to abandon certain of their characteristics by comparing characteristics of those who already waived [19].

10. Electronic Commerce (e-commerce): E-commerce companies used data mining to utilize the knowledge that focuses on mining focuses on analyzing search logs or other activity logs of web users to find interesting patterns. Data mining in e-commerce helps understand customer's wants, love, and value orientation as much as possible to make sure competitiveness in the e-commerce era [20].

Table 1 shows the distribution of the articles according to the application of Data Mining. It shows can be applied to various knowledge resources using different Data Mining algorithms.

Table 1: Distribution of the Articles According to the Application of Data Mining

Authors	Knowledge Resources	Datasets	Data Mining Tasks	Data Mining Algorithms
Yoo et al. (2001)	Healthcare and Biomedicine	Public Health Data • healthcare providers' database • out-patient healthcare statistics database • the medical status database	• classification • clustering • association	• Naïve Bayes Classifier • Neural Network • Decision Tree • Support Vector Machine • Classification based on association (CBA) • Ensemble • AdaBoost • Hierarchical clustering algorithms • Partitional clustering algorithms • KID3 • Apriori
Zhao et al. (2016)	Energy Conservation	Public Buildings • building structure • climate/environmental parameters • construction • system operating condition • user-behavior patterns	• Clustering • Classification	• Chameleon algorithm • genetic algorithm
Alejandrino et al.	Education	Student Academic Data	• Classification	• Naïve Bayes Classifier

(2020)		• Senior High School Academic Data • Demographic Data • Socio-economic data • Admission Test • Tertiary Academic Data		• C4.5 Classifier
Moin et al. (2012)	Banking/Finance	Customer Banking Transactions • Credit • Purchases • stock portfolios • ranking investments	• Classification • Association • Clustering	• Classification by decision tree induction • Bayesian Classification • Neural Networks • Support Vector Machines (SVM) • Classification Based on Association
Liao et al. (2008)	Retail	Customer • Personal data • Historical data • Household data • Purchases Product data • brand	• Association • Clustering	• Decision Tree (C5.0, CART). • Neural Net and RBF Function. • Association Rule (Apriori, GRI). • Sequence Detection. • Clustering Analysis (K-means, Two-step, and Kohonen). • Regression (Linear Regression, Logistic Regression). • Factor Analysis and PCA.
Thuraisingham et al. (2008)	Cybersecurity	• Historical data • Malicious Code • Phone Logs • Web Logs • Transaction Logs	• Classification	• Naïve Bayesian classifiers • Support Vector Machine
Fan et al. (2017)	Collaboration/Team Work	• Open-source software community	• Regression	• Collaboration Process Pattern
Raorane A.A. and Kulkarni R.V. (2013)	Agriculture	• Historical crop yields	• Classification • Regression	•
Arcanu V. (2011)	Customer Relationship Management	• Customers' profile and transaction activities	• Clustering • Regression • Association	• Learning Vector Quantization • Market Basket Analysis
Yadav et al. (2012)	Electronic Commerce (e-commerce)	• Web log files	• Classification • Sequential Pattern • Dependency Modeling • Clustering • Association	• Apriori Algorithm • K-means Algorithm

5.2 Data Mining Techniques/Applications Used in Knowledge Management

The application of data mining has been widely applied in various enterprises and organizations from the context of the articles reviewed. Several business and enterprise domains, ranging from healthcare, education, energy conservation, banking and finance, retailing, cybersecurity, collaboration, customer relationship management, and electronic Commerce. Each sector can be supported by different data mining techniques which generally, include classification, clustering, and dependency modeling. This paper provides a brief description of the four most used data mining techniques, including its standard tools used and some references as follows:

1. Classification

One of the most common data mining tasks is classification. This task maps a data item into one of the various classes [21]. Examples of classification techniques used as part of knowledge management include the classifying of the

patients' diagnosis based on the symptoms because patient symptoms and health conditions are strongly related to each other, and visualization method to facilitate KM and decision-making processes. In the healthcare and biomedicine domain, [12] discussed several classification algorithms such as Naïve Bayesian classifier, Decision tree, Neural network, Support vector machine (SVM), classification based on association (CBA), Ensemble, and AdaBoost. These algorithms are used to correctly identify and predict patients' health conditions based on the symptoms and family health history condition. In the study of [13] in the construction field, classification techniques are used to successfully load prediction, fault detection and diagnosis, and as well as optimal control. Naïve Bayes, decision trees, and support vector machines were applied to generate knowledge in outstanding decision support to facilitate building operational processes. In another study of [14] in the education sector, classification is used to create knowledge to help for the interpretation of student's learning habits, academic performances, and further improvements. Naïve Bayes and decision tree classifiers are used in the prediction of the students 'performance. In the banking and finance industry, a study of [15] discussed classification techniques that have been used in fraud detection and credit risk applications. Classification by decision tree, Bayesian Classification, Neural Networks Support Vector Machines (SVM), and classification Based on associations are the suited algorithms used in fraud detection and credit risk applications.

2. Clustering

Clustering is defined as unsupervised learning that occurs by observing only independent variables, and it involves finding a finite set of categories and grouping objects together that are related to each other and unrelated to the objects belonging to different clusters [21], [12]. This method has been applied in many fields, for example:

- Healthcare and biomedicine: Genetic clustering is a valuable asset when used by researchers who study genes. Clustering has been widely used to study genes when very little information about the specific genes is available and when microarray data can be used for clustering genes [12].
- Banking /Finance: Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as geographic location, job profile, age, gender, interests, spending habits, and so on. This information will help the bank in cross-selling their products and identify which products and services are most relevant to customers.
- Education: Deals with clustering of students' data to predict students' academic behavior and performance and to identify the significant variables that affect and influence the performance of students.
- Retailing: Clustering the segmentation for the probable product line and brand extension to identify the market according to customer clusters [10].
- E-commerce: The clustering analysis is used to classify and group customers and product items with similar characteristics, and it can help with marketing decisions [20].

Conventional data mining algorithms used in clustering are K-means Algorithm, Hierarchical clustering algorithms, partitional clustering algorithms.

3. Association Rule

Frequent itemset/pattern or Association rule mining in the database is often employed to discover customers" hidden sales patterns or relationships among items purchased. It has been widely used in sales transaction databases. An example is if a customer buys bread and butter, then it is most likely the customer buys milk also. In the healthcare and biomedicine field, the association rule is used to discover primary relationships among symptoms, health conditions, and diseases [12]. In the construction industry, association rule mining can be used as an effective decision aid for management to optimize and conserve building operational processes and energy consumption [13]. In the banking industry, association rule mining characterizes the groups of customers by creating customer profiles. Identifying customers using the customer behavior analysis model is a helpful technique to eases the marketing strategy. With this, bank executives can predict with higher accuracy how customers will react to the changes in interest rates, which customers will be likely to accept new product offerings, which consumers will be at a higher risk for defaulting on a loan, and how to make customer relationships more profitable [15]. The most common data mining algorithm used in association rule is the Apriori algorithm.

6. CONCLUSION AND RECOMMENDATION

Knowledge is an important asset in every enterprise and organization. Knowledge management has become a fundamental challenge for development and improvement. Thus, one of the tools to generate knowledge is through the study and analysis of the collection of data or simply referred to as data mining. This paper has identified ten articles on data mining as a tool for knowledge management.

This paper aims to give a research summary of the application of data mining in knowledge management. For several years, data mining has been a great aid in the knowledge management process. Currently, the application of data mining in the knowledge management process in businesses and organizations is still in its infancy. The need to know how to use data mining efficiently to obtain accurate and relevant knowledge is an urgent need. Moreover, this paper recommends that the data used in the knowledge discovery process must be handled with due care and

confidentiality to secure and abide by data privacy, especially in the medical and educational sectors. It is also recommended to use hybrid data mining tasks (e.g., classification and clustering, association rule and clustering, etc.) to sight more knowledge that leads to better performance in decision making.

It is apparent that Web Systems and Technology can have a beneficial effect on the practice of data mining and knowledge management, and that should be an interesting area for future information systems research.

REFERENCES

[1] T. Yigitcanlar, K. Velibeyoglu, and S. Baum, **Knowledge-based urban development: Planning and applications in the information era**. *Information Science Reference (an imprint of IGI Global)*, 2008.

[2] A. Targowski, **From Data to Wisdom**, vol. 15, no. 5. 2005.

[3] J. Han, M. Kamber, and J. Pei, *Data mining concepts and techniques*, 3rd ed. Morgan Kaufmann, 2011.

[4] S. P. Brijesh Kumar Baradwaj, "**Mining Educational Data to Analyze Students" Performance**," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 6, pp. 64–69, 2011, doi: 10.1177/039463200201500108.

[5] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "**From data mining to knowledge discovery in databases,**" *AI Mag.*, vol. 17, no. 3, pp. 37–53, 1996.

[6] F. Gorunescu, *Data mining: concepts, models and techniques*. Springer-Verlag Berlin Heidelberg, 2011.

[7] O. Modinou, L. Liaropoulos, D. Kaitelidou, kyriakos Kioulafas, and E.-M. Theodoraki, **Management of Precancerous Lesions of the Uterine Cervix according to Demographic Data**, vol. 2011. 2011.

[8] M. Gloet and M. Terziovski, "**Exploring the relationship between knowledge management practices and innovation performance,**" *J. Manuf. Technol. Manag.*, vol. 15, no. 5, pp. 402–409, 2004, doi: 10.1108/17410380410540390.

[9] S. Mohapatra, A. Agrawal, and A. Satpathy, "**Designing Knowledge Management-Enabled Business Strategies**," *Manag. Prof.*, pp. 55–88, 2016, doi: 10.1007/978-3-319-33894-1.

[10] S. H. Liao, C. M. Chen, and C. H. Wu, "**Mining customer knowledge for product line and brand extension in retailing,**" *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1763–1776, 2008, doi: 10.1016/j.eswa.2007.01.036.

[11] H. Wang and S. Wang, "**A knowledge management approach to data mining process for business intelligence**," *Ind. Manag. Data Syst.*, vol. 108, no. 5, pp. 622–634, 2008, doi: 10.1108/02635570810876750.

[12] I. Yoo *et al.*, "**Data mining in healthcare and biomedicine: A survey of the literature**," *J. Med. Syst.*, vol. 36, no. 4, pp. 2431–2448, 2012, doi: 10.1007/s10916-011-9710-5.

[13] L. Zhao, J. Zhang, and C. Zhong, "**The Application of Data Mining Technology in Building Energy Consumption Data Analysis,**" *Int. J. Comput. Inf. Eng.*, vol. 10, no. 1, pp. 81–85, 2016.

[14] J. Alejandrino, A. J. Delima, and V. Ramcis, "**IT Students Selection and Admission Analysis using Naïve Bayes and C4 . 5 Algorithm**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, pp. 759–765, 2020, doi: https://doi.org/10.30534/ijatcse/2020/108912020.

[15] K. I. Moin and Q. B. Ahmed, "**Use of Data Mining in Banking,**" *Int. J. Eng. Res. Appl.*, vol. 2, no. 2, pp. 738–742, 2012.

[16] B. Thuraisingham, L. Khan, M. M. Masud, and K. W. Hamlen, "**Data mining for security applications**," in *Proceedings of The 5th International Conference on Embedded and Ubiquitous Computing, EUC 2008*, 2008, vol. 2, pp. 585–589, doi: 10.1109/EUC.2008.62.

[17] S. Fan, X. Li, and J. L. Zhao, "**Collaboration process pattern approach to improving teamwork performance: A data mining-based methodology**," *INFORMS J. Comput.*, vol. 29, no. 3, pp. 438–456, 2017, doi: 10.1287/ijoc.2016.0739.

[18] A. Raorane and R. Kulkarni, "**Review-Role of data mining in Agriculture,**" *Int. J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 270–272, 2013.

[19] V.-B. Arcanu, "**Data Mining In Customer Relationship Management**," *Analele Univ. "Constantin Brâncuşi" din Târgu Jiu, Ser. Econ.*, no. 2, pp. 161–167, 2011.

[20] M. P. Yadav, M. Feeroz, and V. K. Yadav, "**Mining the customer behavior using web usage mining in e-commerce**," *2012 3rd International Conference on Computing, Communication and Networking Technologies*, ICCCNT 2012, 2012, no. July, pp. 0–4, doi: 10.1109/ICCCNT.2012.6395938

[21] T. Silwattananusarn, "**Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012**," *Int. J. Data Min. Knowl. Manag. Process*, vol. 2, no. 5, pp. 13–24, 2012, doi: 10.5121/ijdkp.2012.2502.