Application of Data Mining and Knowledge Management for Business Improvement: An Exploratory Study

Lawal, N.T. A.  
Department of Computer Science, Osun State College Of Technology, Esa-Oke, Nigeria

Denyi, O. A.  
Department of Computer Science, Osun State College Of Technology, Esa-Oke, Nigeria

Kayode, A. A.  
Department of Computer Science, Osun State College Of Technology, Esa-Oke, Nigeria

ABSTRACT

In recent years, there have been a lot of approaches employed by organizations to satisfy their customers and gain competitive advantage. Continuous development of Information System applications is also changing the ways in which businesses are conducted. From scanning barcodes at point of sale (POS) to shopping on the web, businesses are generating large volume of data about products and consumers which are being stored in different data repositories. While a lot of useful knowledge about products, sales and customers that can assist in business decisions are locked away in these databases unexploited. However, the need for organizations to survive in this dynamic business environment depends on how proactive they change these data into useful knowledge which can aid value creation. Presently, customer relationship management and marketing turn out to be the domains which have the potentials to utilize data mining techniques for decision support. This paper examines how business can improve on their performance through utilization of knowledge management (KM) and data mining (DM) applications to manage and support their strategies. Lastly, synergies and challenges of implementation of KM and DM as a tool in business are also critically analysed.

General Terms
Data Mining Techniques, Decision Support.

Keywords
Data, Data Mining, Knowledge, Knowledge Management, Knowledge Discovery in Databases, Business, Marketing

1. INTRODUCTION

At any interval of time, businesses are conducted in all parts of the world. Several private or public corporations are involved in either in production or buying of goods or services, with a view to making profit. All business owners or managers want to make profits in order to broaden the rate of production or sales of goods and services to customers as well as to remain financially healthy in business. Therefore, managers of business whether in private or government or elsewhere must have sufficient knowledge of identification and evaluation of factors that may affect their organization [1].

Business managers like other managers are practical bunch; they are usually not interested in the technology that are used to produce knowledge per se but in the usage of the knowledge produced [2]. Thus, business operation is not all about profitability alone as generally opined but also include among others giving value for money, customer satisfaction, stability and growth of business, innovation, research and development, and marketing opportunities [3].

The main concerns of businesses in the 21st centuries are that companies are finding it even more difficult to predict changes in their competitive environment; customers are becoming competitors as they have alternatives products or service providers, competitors are becoming partners as they offer similar but improved services and better customer relationships, and all these are leading to emerging businesses [4]. However, businesses must go despite scenarios in new business environments that are at present not well understood. Hence, understanding customers’ needs and wants in real time, its translation to value creation in business, integration of innovative IT tools which involves research, design and manufacturing, distribution, marketing and after sale services are some of the concerns of businesses in the 21st centuries. Understanding customer behaviour is also another concern because it has effects on adjustment of business strategies, revenue and identification of new opportunities [5]. Also, the concern on what areas of business would excel also has impact to the point where it might transform the performance of business.

[3] remarked that Information Technology is removing boundaries of time and location, and that demarcations are also fading away between large and small companies. Small, agile firms are now effectively competing with industry giants because IT can make a consortium of small firms look, feel and get big, grasping for customers which are once unreachable by them. [6] opined that as the IT applications proliferate through all business operations, business also increasingly produced vast amount of customer data in database repository which are largely hidden and untapped. A lot of businesses recognize that the knowledge in this huge data repository is significant to support decision making process. Specifically, knowledge about customers is a core heart of business function, however much of this useful data remain untapped. On the other hand, the competition and availability of alternatives for customers as earlier discussed gives new concern to business owners and there has evolved a necessity for creating an effective and efficient customer relationship management. In this circumstance, a knowledge management approach can transform this customer knowledge into business strategies. While the data mining tools and techniques will greatly aid in revealing the hidden and untapped customer data and get to know them better. Knowledge Management (KM) is central to this and is increasingly recognized as an integral part of an organization’s strategy to improve business performance. Therefore, this paper explores how business can improve on
their performance through utilization of knowledge management (KM) and data mining (DM) applications. This is because [7, 8] stated that data mining and knowledge discovery applications have gotten a rich focus due to its significance in decision making. What is more, synergies and challenges of implementation of KM and DM as a tool in business are also part of the study.

The remainder of this paper is organized as follows: The next section presents the concept of knowledge, knowledge management, knowledge management process and knowledge discovery from databases. Section 3 describes the data mining tasks in relation to business scenario. Section 4 surveys the applications of KD and DM in business decision making, and most especially in business domain, while section 5 presents the implementation challenges of KM in business, and the DM application limitations. Finally Section 6 concludes the paper.

2. THE CONCEPT OF KNOWLEDGE AND KNOWLEDGE MANAGEMENT

2.1 Knowledge

Definition of knowledge ranges from practical to the conceptual, to the philosophical, and from narrow to broad in scope. [9] observed that the word ‘knowledge’ means different things to different people. There is no universal description of knowledge; it is only generally agreed that knowledge is a continuum of data and information. Knowledge is a basic and distinct asset that supports the techniques of decision making in business. Knowledge in an organization can exist in structured, semi-structured and unstructured forms as depicted in Table 1.

Ordinarily, data are always in structured form, factual, and often times numeric, and reside in database; Information is factual, but unstructured, and in numerous cases textual; knowledge is inferential, intangible, and is needed to support decision making or postulate a theory [10, 11].

<table>
<thead>
<tr>
<th>Table 1: Different forms of Knowledge Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured</td>
</tr>
<tr>
<td>Financial Data</td>
</tr>
<tr>
<td>Sales Data</td>
</tr>
<tr>
<td>Customer Data</td>
</tr>
<tr>
<td>Demographic Data</td>
</tr>
</tbody>
</table>

[1] Also described knowledge as actionable information which aids individuals to make better decisions and to be more innovative in their methods to a range of problem solving activities. Hitherto, the knowledge has not been properly managed despite its consideration as the substantial factor in the business world, and this might be because there is no understanding of its critical usefulness in business, as well as lack of veritable and guidelines to manage it [12].

From business perspective knowledge can be viewed within the spectrum of implicit (tacit) knowledge and explicit (codified) knowledge; knowledge that resides in people’s memory and which is often difficult to share is implicit or tacit knowledge while knowledge that is stockpiled in an organization’s manuals, procedures, databases, and is thus, easier to share with other persons or parts of an organization is explicit or codified knowledge [12], [9] also opined that the most common distinction of knowledge in the literature is between explicit knowledge, where the information is easy to understand and financially tangible, and tacit knowledge, which is difficult to document or categorise and is non-financially tangible.

2.2 Knowledge Management

Knowledge Management (KM) is about taking data and turning it into useful and appropriate knowledge in a business environment. Knowledge Management has several definitions which results from multidisciplinary nature of KM and the perspectives. [9] observed that there are diverse views of what KM is, and that defining KM has proved extremely difficult due to three factors. The first factor is the intangible nature of knowledge, where ‘knowledge’ itself is an extremely complex concept to define. Second, when the subject being considered is in the management domain the difficulty is compounded even further due to the subjective and eclectic nature of the field. Lastly, when the subject is not only in the management field but is also emerging rather than established, then the difficulty with definitions is even further magnified. [9] reported the various definitions of Knowledge Management as given by different authors. [13] observed that among these many definitions of KM reported by [9], one feature is common, and that is the ability for an organization to share knowledge effectively so as to reach or maintain a competitive advantage. [13] also observed that preservation of knowledge is probably one of the most important functions of an organisation, but not only should organisation be preserving knowledge, however they should also find a way to generate and capture knowledge – this is known as Knowledge Management. [14] viewed KM from human resources perspective as any process of acquiring, capturing, sharing and using knowledge; wherever it resides, to enhance learning and performance in organization, while [15] defined KM as the process of identifying, capturing, organising and disseminating the intellectual assets that are critical to the organisation’s long-term performance.

[1] considered integration of interdisciplinary fields and strategic perspective of KM, and that the KM main objective is to enhance intellectual capital and increase organization’s performance. Therefore, from their definitions of Knowledge Management, KM can be seen as the process formalisation of and access to experience, knowledge, and expertise that provides new know-hows, enable superior performance, promotes innovation, and improve customer value. [11] further stated that “KM is achieving organizational goals through the strategy-driven motivation and facilitation of (knowledge-) workers to develop, enhance and use their capability to interpret data and information (by using available sources of information, experience, skills, culture, character, personality, feelings, and so on) through a process of giving meaning to these data and information”. [16] indicated that several businesses have acknowledged that codifying, sharing and applying knowledge to their business settings will help the organization, and that the growth of knowledge as intangible assets is linked to their competitive approach; the adopted tactic is a direct result of managerial resolutions related to external contexts.

2.3 Knowledge Management Process

Knowledge management process is to gather, categorise, evaluate and share the knowledge with the organization, and that the organization can utilise vital business information to see what needs to be optimized and what is obtainable in the
future. The basic aim of KM is to ensure that the appropriate knowledge is available at the right time in a form that allows timely decision-making [1]. The basic process of knowledge management emphasises on knowledge flows and deals with capture and creation, sharing, and distribution of knowledge. Every unit of knowledge capture and creation, sharing and dissemination, and acquisition and application can be facilitated by knowledge management technologies. Technologies contribute immensely to KM, therefore they are essential apparatus for KM utilization [17]. Hence, technologies provide a platform for KM to enable communications, collaboration and content for better knowledge management process flows. The KM technologies integrated with KM process cycle is as shown in Figure 1.

![Figure 1: KM Technologies integrated with KM Process Cycle (source: [18])](image)

It can be inferred from above figure that knowledge can as well be thought to be to be a cyclical system that enables an organization to efficiently achieve its objectives by having the ability to transform tacit and explicit learnings into habits, better planning, and execution. [18] in their study highlighted six groups in which technologies that enable KM to be fully utilized in business can be categorised as Database Technology, Data Mining, Information and Communication Technology, KM framework, Knowledge-Based System and Modelling.

2.4 Knowledge Discovery from Databases (KDD)

[1] Argued that the development of knowledge discovery in databases in recent years has tightly integrated domain knowledge into discovery process to provide solutions in areas such as marketing, fraud detection, production control and the web. [19] defined Knowledge Discovery from databases (KDD) as: “the non-trivial process of identifying valid, novel, useful and ultimately understandable patterns in the data”. As also proposed by [19], knowledge discovery from databases involves steps in which data mining (DM) is a unique and most significant subfield of extracting knowledge from databases in a novel and patterns for organisation usage. The steps as depicted by [19] are adapted in the Figure 2.

[20] Further described the knowledge discovery process shown in Figure 2 as an iterative structure of the following stages:

1. Data cleaning and integration is the removal of noise, inconsistent data and multiple data sources.
2. Data selection and transformation is retrieval of data relevant to the task from the databases and convert them into types that are appropriate for mining by performing aggregation and summary operations.
3. Data mining is an important process where intelligent methods are applied to extract data patterns.
4. Pattern evaluation is to identify the try interesting patterns representing knowledge based on interesting measures.
5. Knowledge presentation is where visualisation and knowledge representation techniques are used to present mined knowledge to the users.

Stages 1 to 2 are different forms of data pre-processing, where data are prepared for mining while 4 to 5 shows post processing of mined data, while stage 5 shows data mining as one of the phases of KDD.

![Figure 2: Data Mining and KDD Process (Source: [19, 20])](image)

3. DATA MINING (DM)

As earlier noted data mining is a process that allows users to find patterns of knowledge, and uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and subsequent knowledge from large databases. The main functionality of data mining techniques is applying various methods and algorithms in order to discover and extract patterns of stored data [21]. DM tools can answer business questions that traditionally were too time-consuming to resolve. They can search databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectation [20]. The significance of any data mining operation is extraction of innovative patterns which deals with interactions between subsets of data [22].
Data mining tasks can be generally classified as predictive and descriptive; however several authors have used different terms to describe them in broad manner, tasks performed by data mining or types of pattern to be mined. Authors like [6, 19 and 20] outlined data mining tasks using different terms. The tasks as highlighted by [6] are Dependency Analysis, Deviation Detection, Class Identification, Concept Description and Data Visualization. [19] outlined them as Classification, Regression, Summarization, Link Analysis, Dependency Modeling, and Sequence Analysis, while [20] listed these tasks as Concept Description, Association, Classification and Regression For Predictive Analysis, Cluster Analysis, and Outlier Analysis. Although the authors utilized different terms in their studies, the concept is same, except data visualization technique that was only examined by [6]. Hence, this study only discusses the terms description of [6] as related to business scenario in the following paragraphs.

Dependency Analysis: The basic class of dependency analysis knowledge is association between sets of items. This method shows connection between different products that are frequently bought by a customer. As such, this class of knowledge can be deployed in granting discount, conducting promotions for products that have connected relationship, for example Bread and milk are frequently bought together, so a retailer may offer discount on bread if milk and bread are bought together reflecting on the frequency of past purchases that almost all buyers of bread buy milk. Pattern of purchase is bread then milk.

Deviation Detection: This concept deals with identification of a data set that does not conform to general features of the remaining data. This deviant data set is called anomalies or outliers. This is applied mostly in detection of fraud. A credit card industry may notice a change in the withdrawal rates of a customer, different from the usual pattern, though this may be as a result of present need of the cardholder; however it will still be flagged as an outlier to prevent fraud.

Data Visualization: Data visualization process is the display of information in a graphical or tabular format or mapping data to graphical element. Representation of data in graphical forms allows people to interpret visual information easily and find patterns in it. Business owner can easily from graphs or charts quickly identify the fast selling products and buying patterns of their customers. Visualization analysis of product purchases between customers say of Frozen foods in two or more location can be easily understood by sales manager and can aid in some strategic decisions.

Concept Description: Data entries such as customers can be grouped with classes, concepts or domain knowledge. This method covers summarization or characterization, discrimination and comparison data of a class. Data summarization is a process of obtaining a general characteristic summary of a data of a class under study (target class) while data discrimination is a comparison of the target class with one or set of comparative class. For instance, for data summarization, to study the characteristics of say a customer in a retail stores whose purchases drops by 20% in April, the data related to such customer can be collected by using a single database query say SQL; and for data discrimination, to study a customer whose consumption is 10% greater than customer with worst purchase in April 2014. The same method applies to extract the customer information but the resulting output in graphs curves, charts will differ as discrimination concept will have discriminant rules [20]. This will allow the marketing department to know if a customer is dropping its loyalty or changing taste by profiling their data.

Clustering: This is the task of segmenting heterogeneous items into a number of more homogenous subgroups or clusters. Records with same features are classified together, the knowledge used decide their meaning, if any, to attach to the resulting groupings. For instance, clustering may be the first step in market segmentation. Instead of thinking on what kind of meat does student may prefer to be discounted by a supplier, supplier's marketing department must group students on religious and cultural basis or their previous buying patterns and then decide which meat to go for promotion.

4. APPLICATIONS OF DM AND KM

4.1 DM and KM Applications in Business Decision Making

Applications of DM to KM give immense supports to business decision making, and the resultant effect of the interaction creates business intelligence (BI). An illustration of their interactions to generate BI is shown in Figure 3. BI is the ability of a business to correctly transform and utilize information and knowledge in a timely manner for business development [23].

From the illustration, it can be easily formulated that data mining techniques are able to find hidden and unknown information for organisation to gain knowledge, which results in provision of BI. With the creation of BI, business managers will find it easy to answer some of the following questions: What is the frequency of purchase of a product by customers? The knowledge from customer profile will help marketing department formulates their decision on product to promote for whose customers.

4.2 DM and KM Applications in Business Domains

In this section, few application domains of DM and KM are briefly examined. Areas of business where DM and KM are majorly used are retailing, banking, telecommunication, insurance, and manufacturing to mention a few. All these organisations use data mining tools with other techniques such as statistics, pattern recognition and other important tools, to find patterns and connections, in form of knowledge that otherwise would have been hard to locate, for critical decision making in their businesses.

1. Retailing: This is most fit industrial area of DM for KM, as it gathers vast amount of data on sales, customer shopping history, point of sales reading several barcodes, etc. In fact, the data collected at points of sale in a moment by Wal-Mart, US largest retailer, is in terabytes [1]. The quantity of data gathered continues to
escalate due to the increasing availability, ease, and popularity of e-commerce. As at this moment, some chain stores have websites where customers can buy services or products, for instance, amazon.com exists solely online without any physical stores. Frankly speaking, retail is a rich source for mining data to build knowledge to identify customers buy behavior and trends, achieve better customer retention and satisfaction, improve good consumption ratio, and distribution policies and ultimately reduced the operational cost of business [20]. Data mining techniques such as association can facilitate discovery of buying patterns like a customer that buy bread will buy milk scenarios and deviation analysis can help a retail store to identify a potential fraudsters. Consumer profiling also helps significantly.

2. Banking and Finance: This sector has significantly benefitted equally from the advent of DM and KM. Anywhere in the world, financial services use databases to store a huge amount of data DM can contribute to solving business problems in banking and finance by finding patterns, correlation, deviation in business information and market prices that are not apparent to managers as data is too voluminous for human processing. DM and KM are interest of financial services because they offer them the potential to reorganise their global risk and return on investment [24]. DM and KM has greatly assisted financial businesses in the areas of managing risk, trading and customer relationship, portfolio, as well as marketing strategies. For instance, in a loan offer, banks can explore past customers loan repayment patterns. Fraud is a great concern, in its detection, outlier or deviation analysis of past transactions can identify patterns that fall out of general characteristic which be a likely fraud.

3. Aviation: This sector if properly manage, its customer knowledge can easily be identify by association rule or clustering technique, customers that can be given discounts to fly more by detecting flying frequency and class of flight.

4. Insurance companies: They can also easily get in its net more customers via effectiveness of its campaigns, optimization and execution, decreased mailing costs and increase conversion rates, forecasting which customer will buy new policies and how much.

5. Manufacturing: Products are now being manufactured to specification on customers’ requests through choice board and thereby allows manufacturer to know which product to produce for who and how. Therefore manufacturing companies must be able to use its knowledge based mechanism to predict which features should be bundled to meet customer demand. In case of warranties for accompany products, manufacturers can as well predict the amount of customers that will return products as a result of failure, submit warranty claims and the likely sum of the claims.

6. Online Business: Major online companies and catalogues such as amazon.com, betterware.co.uk, and jumia.com are using the information stored in customer profile to offer personalized products recommendations based on opinion of other customers. Several of these online companies use personalized store for every customer and use DM techniques to identify customers’ interest and make reliable recommendations [22].

7. Health Care: The health industries can correlate using clustering method, the demographics of patient with serious ailment say HIV and cancer, and doctor having this knowledge can develop better perception of symptoms and how to provide better diagnosis tool and therapy [4].

8. Broadcasting: The capability to know which programmes are featured in a peak period can assist in predicting when to place advertisement.

5. IMPLEMENTATION CHALLENGES AND LIMITATIONS OF KM AND DM

5.1 Implementation Challenges of KM in Business

There are associated challenges with implementation of KM in a business process, the need of knowledge management in business is not doubtful, however, various enterprises still face numerous encounters in applying a workable and actual process for knowledge management in business [4]. [25] identified the following challenges:

1. Technology: Integration of information system applications with business process is a critical challenge due to complacency of the people with the old system. Change is inevitable; therefore knowledge management is not precluded to what an organization knows but related to aggregation of information system with people with a view to getting overall capabilities of knowledge management in business [25].

2. Individual: The learning of skills and mindset of individual which result in tacit knowledge is another serious challenge to group idea of sharing knowledge. Since tacit knowledge are difficult to share unless articulated. However, emphasis must be placed on explicit form of knowledge, and an approach allowing tacit knowledge sharing must be supported by management to surmount these challenges. Motivation for knowledge workers willing to share his tacit knowledge can be a catalyst for knowledge sharing [25].

3. Shared leadership: Some organizations do not give knowledge workers leadership responsibilities which make the knowledge workers to feel indifferent to whatever may be the reason of top management to implement KM in the organization [25]. Lack of shared leadership can pose a threat to a successful implementation of KM in business.

4. People and Culture: Knowledge and experiences needed by business resides in human and not in technologies, technologies can only speedily capture information or knowledge and process them [25]. So cultural concept must be developed to facilitate sharing and interpersonal relation to make KM implementation succeeds.

4.1 Data Mining Applications Limitations

The major issue in data mining applications limitations is privacy and data use. The technology makes it possible to analyze routine business transactions and gets substantial information about individuals buying behaviour and
preferences. Unless data protection act of individual is reviewed, it may not be so ethical for analyzing people’s profiles without their consent despite the fact that the mining is for their satisfaction.

There is also a serious discussion on how data must be represented in databases in one format to remove heterogeneous data representation in tables, arrays, cubes, and so on. Data reliability is another issue as there is a great need to integrate conflicting data of the same record from different repositories, for instance customer full names may be used in a particular profiling while in another it is family and initials. Hardware price have fallen considerably in recent years, data mining and other associated technology price is escalating, but vast amount of data are being accumulated, data mining tools are needed but the rising cost is an issue. In addition, from the technology point of view scalability, dimensionality, data streaming, mining from different sources, iterative mining from multiple abstraction and integration of techniques together as a time that is clustering and association simultaneously are still challenges in data mining applications.

5. CONCLUSIONS
This paper has examined the various areas of businesses challenges, where data mining can greatly assist in discovery, extraction, creation, usage of knowledge patterns for KM, which ordinarily are wasting away in various storage media from business activities. It has also been found that KM is a framework that all organizations must adequately utilize to its maxim if competitive advantage is a mission to accomplish. Customers are now gods of businesses, thus all managerial efforts to meet their tastes, demands, preferences and satisfaction deserve a serious commitment. Data mining tools are the catalysts that businesses can use to attain their objectives of profitability, competitive advantage and overall, building a functional customer relationship management and effective marketing strategies.

Businesses, however, as noted from the review still have not fully utilize the data that are streaming every second as technology advancement has also escalated data capture and accumulation abilities. Though, there are several challenges which implementation of KM may face in an organization, its usage in all business in the nearest future is inevitable for an enterprise that wants to sail through the competitive markets. More so, these challenges are identifiable and can be solved systematically. There is also a need to fuse various techniques of data mining for better knowledge creation and see how some other innovative approaches can emerge. Indeed, they are sure to emerge. It is clear that the data mining techniques will have a major impact on the practice of KM, and will present significant challenges for the future of business processes. Finally, the call for ubiquitous data mining, web mining and other mining tools are needed now than ever as data accumulation will continue to be on increase as part of our lives. DM is a veritable tool for KM, just like any other tool. It is not sufficient to know how DM works, but to know how it is used. So, DM suggests KM, but businesses decide KM.

6. REFERENCES


