# ROLE OF KNOWLEDGE MANAGEMENT TECHNIQUES IN DATA MINING FOR LARGE SCALE DATA IN BANKS

**Renu Chauhan[1], Dr. Rajeev Kumar[2]**

**Department of Computer Science and Engineering**

**[1,2]Shri Venkateshwara University, Gajraula (Uttar Pradesh)**

## Abstract

In recent years Knowledge Management (KM) has become an important contributor to the success of enterprises. Increasing product complexity, globalization, virtual organizations and customer-oriented demand require a more thorough and systematic management of knowledge both within individual banks and between several cooperating banks. Information Technology (IT) supported KM solutions are often built around an organizational memory that integrates informal, semi-formal and formal knowledge to facilitate the access, sharing and reuse of knowledge by members of the banks to solve their individual or collective tasks. In such a context, knowledge has to be modeled, appropriately structured and interlinked to support flexible integration and its personalized presentation to the customer. Knowledge discovery and knowledge management solutions can provide great benefits and better services in many banks contexts. Business Intelligence Systems may be used to help to uncover useful information and convert it into a form of knowledge by utilizing Data and Text Mining techniques in banking. Data mining techniques are a valuable help for the banking system that is related to better banking growth by providing new customers.

## 1. OVERVIEW

Today banking systems collect the large amount of data on a daily basis. Therefore, the collected data are customer information, transaction details and credit card details. Many decisions are made in a day. Decisions taken include the start of the relationship, investment decisions, credit decisions, predetermined decisions. Banking systems depend on many decisions and must be corrected automatically. But this is a manual process and is error prone and time consuming due to the large volume of transactional and historical data. In this article, explore and review data mining techniques in the banking sector. The key DM techniques adopted in the banking sector are identified, which include cluster analysis, association rule extraction (ARM) and classification techniques, including, among others, decision trees (DTs), networks neural (NN), support vector machines (SVM), Naive Bayes (NB) and Logistic Regression (LR).

Data mining techniques are a valuable help for the banking system that is related to better banking growth by providing new customers and treating old customers

through the use of some different techniques, such as completing goals within the given time period for banks and other related systems, how to combine the required activity in the single cluster known as clustering so that they can be easily performed, providing segmented products for better customer selection, rapid fraud detection, model analysis Customer acquisition in the time for appropriate retention and reporting, detection of emerging trends to take better care in a current competitive market that adds much more value to existing products and services and launches new product and service packages. Data mining has a very large application domain in almost all fields where data is generated, so data mining is considered one of the most useful tools in the database and information systems and one of the most effective multidisciplinary developments in the field of the system banking.

## 2. DATA MINING

Data mining techniques have influenced various sectors of life ranging from finances to life sciences but these techniques have only been used for exploiting manufacturing data for discovery of useful knowledge in the last decade or two. Handling these data and information sources usually requires additional knowledge to be provided by domain experts. The reported tools such as clustering and association rules analysis have been tested in manufacturing related data analysis, so the implementation of these tools has been tested along with their scope for discovering useful knowledge by exploiting term based relationships and the refinement of this process has also been considered [1]. Finally, some supervised classification techniques are used to classify data into two classes in terms of good or bad information documents.

On a basic level, data mining is not specific to one kind of media or data. Data mining ought to be relevant to any sort of information storehouse. Be that as it may, algorithms and approaches may contrast when connected to various sorts of data. For sure, the difficulties introduced by various sorts of data differ altogether. Data mining is being put into utilization and concentrated for databases, including relational databases, question relational databases and protest situated databases, data stockrooms, value-based databases, unstructured and semi-organized archives, for example, the World Wide Web, propelled databases, for example, spatial databases, sight and sound databases, time-arrangement databases and textual databases, and even level documents. Here are a few cases in more detail [2]:

➢ **Flat files:** Flat records are really the most widely recognized data hotspot for data mining algorithms, particularly at the examination level. Level records are straightforward data documents in text or paired arrangement with a structure known by the data mining calculation to be connected.

➢ **Relational Databases:** Briefly, a relational database comprises of a set of tables containing either estimations of element attributes, or estimations of attributes from element relationships. Tables have segments and lines, where segments speak to attributes and lines speak to tuples. A tuples in a relational table compares to either a protest or a connection amongst objects and is recognized by a set of quality esteems speaking to a one of a kind key. In Figure 1 we exhibit a few relations Customer, Items, and Borrow speaking to business action in an invented video store OurVideoStore. These relations are only a subset of what could be a database for the video store and is given for instance.
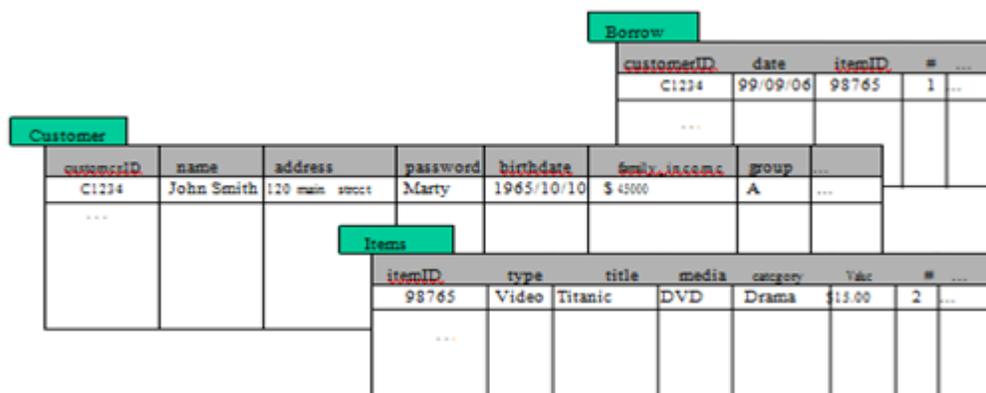


**Figure 1: Fragments of some relations from a relation for our video store**

➢ **Data Warehouses:** A data warehouse as a storage facility is a vault of data gathered from different data sources (often heterogeneous) and is expected to be utilized all in all under the same bound together mapping. A data warehouse gives the choice to examine data from various sources under a similar roof. Give us a chance to assume that OurVideoStore turns into an establishment in North America. Numerous video stores having a place with OurVideoStore organization may have diverse databases and distinctive structures [3]. On the off chance that the official of the organization needs to get to the data from all stores for strategic decision-making, future direction, marketing, and so forth, it would be more appropriate to store every one of the data in one site with a homogeneous structure that permits intelligent analysis. As it were, data from the distinctive stores would be stacked, cleaned, changed and integrated together. Figure 1.2 demonstrates a case of a three-dimensional subset of a data block structure utilized for OurVideoStore data warehouse.
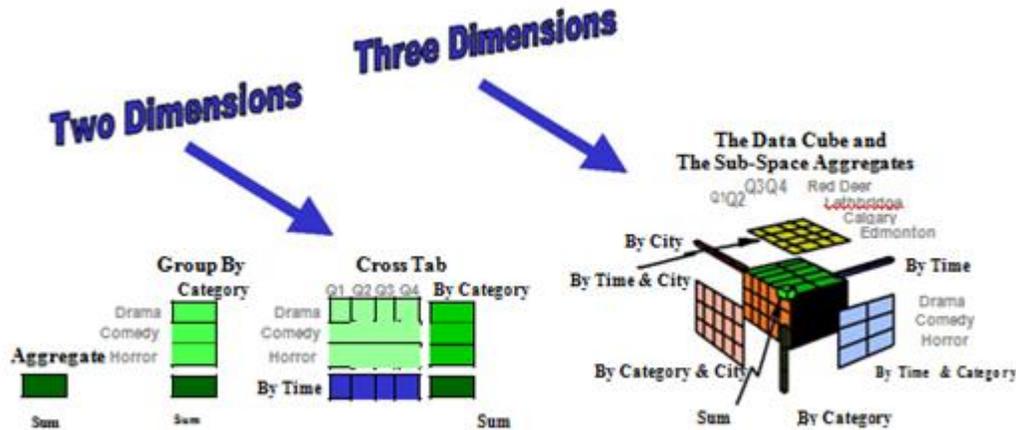
**Figure 2: A multidimensional data cube structure commonly used in data for data warehousing**
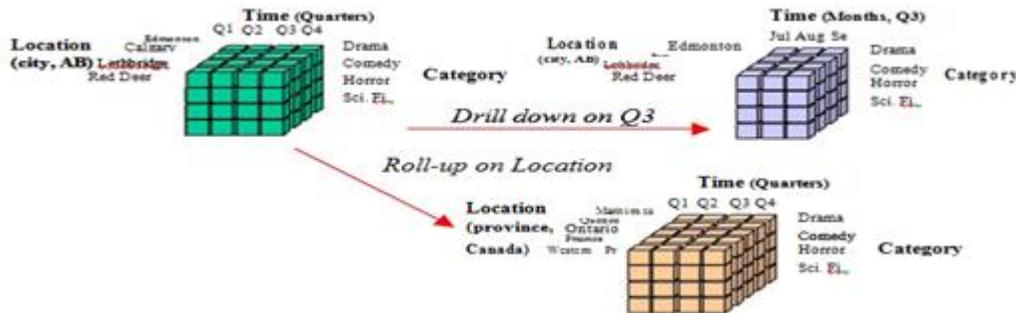


**Figure 3: summarized data from OurVideoStore before and after drill down and roll up operations**

➢ **Transaction Databases:** A transaction database is a set of records speaking to transactions, each with a period stamp, an identifier and a set of things. Related with the transaction documents could likewise be descriptive data for the things [4]. For instance, on account of the video store, the rentals table, for example, appeared in Figure 1.4, speaks to the transaction database. Each record is a rental contract with a client identifier, a date, and the rundown of things leased (i.e. video tapes, recreations, VCR, and so on.). Since relational databases don't permit settled tables (i.e. a set as property estimation), transactions are typically put away in level documents or put away in two standardized transaction tables, one for the transactions and one for the transaction things.

**Figure 4: Fragment of a transaction database for the rentals at Our Video Store**

➤ **Multimedia Databases**: Multimedia databases incorporate video, pictures, and sound and text media. They can be put away on broadened protest relational or question situated databases, or just on a record framework. Multimedia is described by its high dimensional, which makes data mining significantly all the more difficult. Data mining from multimedia stores may require PC vision, PC designs, picture interpretation, and normal dialect processing systems.

➤ **Spatial Databases**: Spatial databases will be databases that, notwithstanding regular data, store topographical information like maps, and worldwide or local situating [5]. Such spatial databases introduce new difficulties to data mining algorithms.
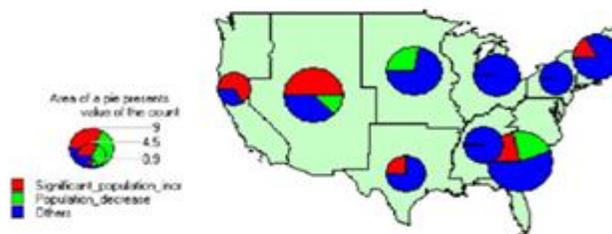


**Figure 5: visualization of spatial OLAP**

➤ **World Wide Web:** The World Wide Web is the most heterogeneous and dynamic archive accessible. An extensive number of creators and distributers are consistently adding to its development and transformation, and countless are getting to its resources every day. Data in the World Wide Web is composed in between associated reports. These archives can be text, sound, video, raw data, and even applications. Theoretically, the World Wide Web is included three noteworthy segments: The substance of the Web, which envelops reports accessible; the structure of the Web, which covers the hyperlinks and the relationships amongst archives; and the utilization of the web, portraying how and when the resources are gotten too. A fourth measurement can be included relating the dynamic nature or advancement of the archives.

Data mining in the World Wide Web, or web mining, tries to address every one of these issues and is often separated into web content mining, web structure mining and web use mining.

## 3. APPLICATIONS OF DATA MINING IN BANKING

Data mining has been utilized broadly by numerous associations. Huge numbers of these associations are joining data mining with things, for example, insights, design acknowledgment, and other imperative devices. Data mining can be utilized to discover examples and associations that would somehow, or another be hard to discover [6].

The banking and finance industry are fertile ground for Data Mining. Banks and financial institutions generate large volumes of detailed transaction data. Fraud detection, risk assessment of potential customers, trend analysis, and direct marketing are the primary Data Mining applications at banks.

In the financial area, requirements for forecasting dominate. Forecasting of stock prices and commodity prices with a high level of approximation can mean large profits. Neural network algorithms are used in forecasting, options and bond trading, portfolio management, and in mergers and acquisitions.

A few examples of Data Mining in the banking and finance are outlined as follows:

- Loan payment prediction and customer credit policy analysis - Loan payment prediction and customer credit analysis are critical to the business of a bank. Many factors can strongly or weakly influence loan payment performance and customer credit rating. Data Mining methods, such as attribute selection and attribute relevance ranking, may help identify important factors and eliminate irrelevant ones.

- Classification and clustering of customers for targeted marketing - Classification and clustering methods can be used for customer group identification and targeted marketing. Customers with similar behaviors regarding loan payments may be identified by multidimensional clustering techniques. These can help identify customer groups, associate a new customer with an appropriate customer group, and facilitate targeted marketing.

- Detection of money laundering and other financial crimes - To detect money laundering and other financial crimes, it is important to integrate information from multiple databases (like bank transaction databases, and federal or state crime history databases), as long as they are potentially related to the study. Multiple data analysis tools can then be used to detect unusual patterns, such as large amounts of cash flow at certain periods, by certain groups of customers.

## 4. KNOWLEDGE MANAGEMENT IN

## BANKS

### *Definition of knowledge*

Defining the meaning of knowledge is and has for long been a subject occupying the minds of theorists and researchers. The word can be used in many different areas and with various notions. In the field of IT and IS, knowledge can be addressed in relation to data and information. The understanding of the relationship between these three can be described in this way: data is raw numbers and raw facts, information on the other hand can be regarded as processed data, and knowledge in this sense is authenticated information. This way of viewing knowledge is still not a completely clear definition. To effectively distinguishing information and knowledge is a problem when it comes to content, accuracy and structure. To make it easier, it is possible to say that knowledge is the information that is being processed on an individual level. In this sense knowledge varies from the different personal views in relation to facts, procedures, concepts, interpretations, ideas, observations, and judgments [7]. Through these various perspectives knowledge can be de-scribed as personalized information, which is affected by the background and experience of the holder

The study develops the knowledge of industrial resources in three banks investigated, such as industrial associations, competitors, customers and suppliers. It has shown that banks are adopting the evolving environment and may be more proactive than reactive.

KM can help banks to adopt a well-defined placement policy to ensure that the right person is in the right place to meet the growing challenges. Frequent job rotation also contributes to the creation of internal knowledge.

1. In this global age, the positive side of any organization is the younger generation, which with its current higher ratings, is generally prepared to take on more responsibilities. What is required is a lot of mentoring by experienced and successful superiors. The experience shared by senior / qualified employees will have a greater impact on this matter.

2. Banks must continually create new products and value for their customers, as their preferences have diversified more in recent years. It is important that banks monitor these preferences and relevant information. In other words, banks must make the best use of their clients' information and the intrinsic knowledge of their employees. Therefore, banks must acquire knowledge from internal and external sources of knowledge.

3. The ability to learn throughout the group and form an effective team for a specific project / task is a powerful tool. KM will provide a construction and banking environment by promoting

collaborations and push mechanisms. KM will also offer a general discussion forum, where employees can share / discuss / talk about economics, business, finance and other related topics.

KM will substantially enhance the productivity of individuals and groups in a bank by allowing to use and reuse of knowledge and doing strategic tasks such as customer relation management.

## 5. DATA MINING AND KNOWLEDGEMANAGEMENT IN BANKS

The banking sector is always aimed at improving customer satisfaction, which will result in increased revenues. The process of creating, archiving and dispersing knowledge becomes essential and banks assign specialized personnel to monitor and manage these critical processes. The most common fields of knowledge management applications in a bank are risk management, marketing management, customer relationship management and performance measurement, particularly for the benefit of its stakeholders. In general, investments in major banks in knowledge management systems, such as the Decision Support System, Data Warehouses and Fata mining are growing rapidly. The bank managers of these days use the computerized support system as personal support in the decision-making process. The knowledge management system in banks guarantees

better and more efficient results in the decision-making process[8].

## 6. CONCLUSION

This research shows data mining applications in banking sectors. It contains a general description of data mining, which provides a definition of the concept, primary mining techniques, and mentions the main fields in which data mining can be applied. It also presents the banking sector that can benefit from the use of DM tools, along with their use cases, i.e. the retail banking sector and the insurance sector. Data has grown in terms of dimensionality and size. With the progress of data mining and know-how, this mountain of data is becoming the most valuable asset of the organization. Precious knowledge and interesting models are hidden in this data. There is great potential for banks to apply data mining in their decision-making processes in areas such as marketing, credit risk management, money laundering detection, liquidity management, investment banking and fraudulent transaction detection in time for the bank such as loss of customers for competition, financial loss, loss of reputation and heavy fines by regulators.

Hence it is concluded that the research reported in this thesis examines ways of discovering useful knowledge through the classification of textual data into one of two categories of document i.e. good or bad information classes. The system proposed is based on Hybrid Applications of Data and

Text Mining Techniques to provide knowledge discovery and management solutions for industrial or business environments. The term "*document*" used in this research work refers to the set of information available in the form of textual data under different headings (i.e. time, cost, and planning) in Post Project Reviews (PPRs).

## REFERENCES

[1]. Pohle, C. (2003). Integrating and Updating Domain Knowledge with Data Mining VLDB, Ph.D. Workshop.

[2]. Manabendra, N, & Choudhury, K. (2009). Exploring the Dimensionality of Service Quality: an Application of TOPSIS in the Indian Banking Industry. APJOR 26(1):115-13.

[3]. MaimonO.andRokach L., (2010). "Data Mining and Knowledge Discovery Handbook", Second Edition, Springer Publishing Company, USA,.

[4]. Cunha, D., Agard, B., and Kusiak, A. (2006). "Data Mining for Improvement of Product Quality." International Journal of Production Research **44**(18-19): 4027-4041.

[5]. Lin, F., Huang, K. and Chen, N. (2005). "Integrating Information Retrieval and Data Mining to Discover Project Team Co-ordination Patterns." Decision Support Systems.

[6]. Agard, B. and Kusiak, A. (2004a). "Data Mining-based Methodology for Product Families." International Journal of Production Research **42**(15): 2955-2969.

[7]. Brezocnik, M., Balic, J. and Brezocnik, Z. (2003). "Emergence of Intelligence in Next Generation Manufacturing Systems." Robotics Computer Integrated Manufacturing **19**: 55-63.

[8]. Mansingh, G.; Rao, L.; Osei-Bryson, K.M.; Mills, A. (2015). Profiling internet banking users: A knowledge discovery in data mining process model based approach. Inf. Syst. Front. 17:193–215.