THE EFFECTIVENESS OF DATA MINING TECHNIQUES IN BANKING

Yuvika Priyadarshini

Researcher, Jharkhand Rai University, Ranchi.

ABSTRACT

The aim of this study is to identify the extent of Data mining activities that are practiced by banks, Data mining is the ability to link structured and unstructured information with the changing rules by which people apply it. It is not a technology, but a solution that applies information technologies. Currently several industries including like banking, finance, retail, insurance, publicity, database marketing, sales predict, etc are Data Mining tools for Customer . Leading banks are using Data Mining tools for customer segmentation and benefit, credit scoring and approval, predicting payment lapse, marketing, detecting illegal transactions, etc. The Banking is realizing that it is possible to gain competitive advantage deploy data mining. This article provides the effectiveness of Data mining technique in organized Banking. It also discusses standard tasks involved in data mining; evaluate various data mining applications in different sectors

KEYWORDS

Definition of Data Mining and its task, Effectiveness of Data Mining Technique, Application of Data Mining in Banking, Global Banking Industry Trends, Effective Data Mining Component and Capabilities, Data Mining Strategy, Benefit of Data Mining Program in Banking

1. DEFINITION OF DATA MINING AND ITS TASK

Data mining is an essential step in the knowledge discovery in databases (KDD) process that produces useful patterns. The terms of KDD and data mining are different. KDD refers to the overall process of discovering useful knowledge from data. Data mining refers to discover new patterns from a wealth of data in databases to extract useful knowledge**.** KDD process consists of iterative sequence methods as follows:

1. *Selection*: Selecting data relevant to the analysis task from the database

2. *Preprocessing*: Removing noise and inconsistent data; combining multiple data sources

3. *Transformation*: Transforming data into appropriate forms to perform data mining

4. *Data mining*: Choosing a data mining algorithm which is appropriate to pattern in the data; Extracting data patterns

5. *Interpretation/Evaluation* : Interpreting the patterns into knowledge by removing redundant or irrelevant patterns; Translating the useful patterns into terms that human understandable

Data mining has two primary objectives of prediction and description. Prediction involves using some variables in data sets in order to predict unknown values of other relevant variables (*e.g.classification, regression, and anomaly detection*).

Data mining task define six main functions of Data mining

1. *Classification* is finding models that analyze and classify a data item into several predefined classes

2. *Regression* is mapping a data item to a real-valued prediction variable

3. *Clustering* is identifying a finite set of categories or clusters to describe the data

4. *Dependency Modeling* (*Association Rule Learning*) is finding a model which describes significant dependencies between variables

5. *Deviation Detection* (*Anomaly Detection*) is discovering the most significant changes in the data

6. *Summarization* is finding a compact description for a subset of data

2. EFFECTIVENESS OF DATA MINING TECHNIQUE

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. DM techniques are the result of a long process of research and product development.

Data mining consists of five major elements; to extract, to transform, and to load transaction data onto the data warehouse system, to store and manage the data in a multidimensional database system, to provide data access to business analysts and in format Analyze the data by application software, and finally to present the data in a useful format, such as a graph or table

DM techniques usually fall into two categories, predictive or descriptive. Predictive DM uses historical data to infer something about future events. Predictive mining tasks use data to build a model to make predictions on unseen future events. Descriptive DM aims to find patterns in the data that provide some information about internal hidden relationships.

Descriptive mining tasks characterize the general properties of the data and represent it in a meaningful way.

Data sources of a data mining system can be divergent information repositories like database, data warehouse or other repository. Data mining engine is the core of data mining system. The functional modules od Data mining algorithms and rules are kept in the engine. Database stores the knowledge that is used to guide the data mining process and provides the data that pattern evaluation module needs to validate the result of data mining.

3. APPLICATION OF DATA MINING IN BANKING

Data Mining can help by contributing in solving business problems by finding patterns, associations and correlations which are hidden in the business information stored in the data bases The banks who have realized the importance of data mining are in the process of reaping huge

profits and considerable competitive advantage. According to the regulations given by Reserve Bank of India, the banks have to Provide Off-site Monitoring Surveillance (OSMOS) reports on regular basis in electronic format only and Regulatory requirement of filing of statutory returns such as the one under Section 42 of the Reserve Bank of India Act, 1934 for working out Cash Reserve Ratio (CRR) and Statutory Liquidity Ratio (SLR) obligations in electronic format. According to the Committee formed by Reserve Bank of India Headed by Dr. A. Vasudevan to go through the details of this topic, gave his report on 17th July,1999, the committee highlighted that by the use of data mining techniques, data available at various computer systems can be accessed and by a combination of techniques like classification, clustering, segmentation, association rules, sequencing, decision tree various ALM reports such as Statement of Structural Liquidity, Statement of Interest Rate Sensitivity etc. or accounting reports like Balance Sheet and Profit & Loss Account can be generated instantaneously for any desired period/ date . Trends can be analyzed and predicted with the availability of historical data and the data warehouse assures that everyone is using the same data at the same level of extraction, which eliminates conflicting analytical results and arguments over the source and quality of data used for analysis. In short, data warehouse enables information processing to be done in a credible, efficient manner. The Committee recognizes the need for data warehouses and data mining both at the individual bank level and at industry level . The implication of adopting such technology in a bank would be as under :

• All transactions captured at the branch level would get consolidated at a central location. Such a central location could be called the Data Warehouse of the concerned bank. For this to happen, one of the requirements would be to establish connectivity between the branches on the one hand and the Data Warehouse platform on the other.

• For banks with large number of branches, it may not be desirable to consolidate the transaction details at one place only. It can be decentralized

• By way of data mining techniques, data available at various computer systems can be accessed and by a combination.

4. GLOBAL BANKING INDUSTRY TRENDS

The ongoing global financial crisis, with its historic dimensions, will have a lasting impact on the global banking industry and the world economy. Banks are looking for growth opportunities2, but their success is very much dependent on their ability to build critical mass and successful operations in these economic times.

The regulatory landscape has strengthened significantly, with governments in many markets implementing much more stringent rules—such as minimum capital requirements—putting pressure on firms to raise capital.

This financial pressure has increased focus on operational efficiencies and is driving investments in automating credit underwriting platforms with scorecards and models, fraud and collection analytics, and enhancing data and risk analytics capabilities. Some banks are progressing from using qualitative or analytical models to the deployment of predictive models for risk and customer analytics.

Amid the increased requirements on capital adequacy and optimal utilization of capital, economic capital management and risk adjusted return on capital (RAROC) have become a top priority for banks. Loss forecasting and stress testing have gained increased importance in the current economic scenario and are the norm today.Other important trends in the global banking industry are:

■ **Business intelligence** practices are being integrated with customer relation management (CRM) and there is a stronger focus on predictive and behavioral modeling tools. This combination is fast becoming the central cog for cross selling, customer lifetime value management, risk management and recovery management.

■ Banks have started to realize the potential of **social media analytics and content tracking** and are aligning their CRM tools with Twitter and Facebook.

■ With the growth of fraudsters and hackers, **security threats** for all firms but especially banks have mushroomed. Among the big drivers of payments security are new encryption initiatives and efforts to bring interoperability to point-to-point transaction encryption.3

■ With the emergence of **mobile banking**, banks have to restructure their business model and increase collaboration with firms in cards, payments and telecom domains.

■ Banks will be ramping up their **personal finance management** (PFMs) offerings to help customers realize their financial goals and to avoid disintermediation

5. EFFECTIVE DATA MINING COMPONENT AND CAPABILITIES

A comprehensive EDM program comprises of a host of capabilities. In order to enable these capabilities, there are a few components that first need to be in place. The pre-requisite components and key capabilities are listed below.

5.1 PRE-REQUISITE COMPONENTS

■ DATA MANAGEMENT VISION – an organization needs to describe the vision and principles or core values around which its enterprise data management program is based.
■ DATA MANAGEMENT GOALS – goals of an EDM program need to be related to strategic business goals, objectives and priorities. These, furthermore, need to be adopted by and communicated to key stakeholders.

■ GOVERNANCE MODEL – An EDM program needs to adopt an enterprise-wide mechanism by which the program is managed, funded and implemented.

■ ISSUES MANAGEMENT AND RESOLUTION – the organization has the ability to identify, triage, track, and update status for all data and integration issues identified during "business as usual" (BAU) activities or ongoing data management initiatives.

■ MONITORING AND CONTROL – Collective capabilities for measuring and reporting on the quality and effectiveness of the data management program as it operates as part of the BAU environment.

5.2 DATA MINING CAPABILITIES

■ **CRITICAL DATA INVENTORY** – Critical Data consists of those data elements (and their business definitions) that the business deems as important for decision making and compliance. This inventory should be made in consultation with business users. The Critical Data Inventory helps prioritize or contain the scope of an EDM program.

■ **DATA INTEGRATION** – This covers the processes and tools for acquisition, composition and enrichment of data from different sources into a single unified store or view. Data integration typically is done by building an enterprise data warehouse, from which data is sourced directly into analytical engines, or into data marts that feed the analytical engines. Data integration also addresses the control processes that are used to monitor data integrity as data flows from data producers to data consumers.

■ **DATA PROFILING** – Data profiling is the examination of data to collect statistics and characteristics about the structure of available data. It is used to assist in critical data assessment, data classification, data integration and impact analysis.

■ **DATA QUALITY** – Data quality measures whether data is 'fit for intended use'. Data quality is typically measured along the dimensions of accuracy, completeness, conformity, consistency, duplication and integrity, with each dimension carrying different weight based on the intended use of the data. End-to-end data quality allows for comparison of data quality across the data flow at a point in time as well as across time (trends).

■ **METADATA MANAGEMENT** – Metadata is information about the data itself. Metadata captures attributes of data like the type, length, timestamp, source, owner etc., as well as relationships in data (semantics), and helps with data traceability and lineage. Use of uniform methods and tools for defining, collecting, and managing information metadata ensures that data is identified consistently across the enterprise.

■ **MASTER DATA MANAGEMENT** – Master data or Master file is the single, authoritative and agreed upon source of data that is critical for business operation. It typically includes persistent non-transactional data like customer, product, employee etc. Master data management ensures that there is a single consistent version of critical data used across the enterprise.

■ **REFERENCE DATA MANAGEMENT** – Reference data is used to classify or categorize data. An example is the product master which contains the list of all products along with their attributes. As with metadata and master data, reference data management also plays an important role in data integrity and consistency.

■ **DATA PRIVACY (ANONYMIZATION)** – This includes processes, algorithms and technology platforms which are required to ensure that the contents of any information object (data set) fully comply with information privacy and protection laws and regulations.

6. DATA MINING STRATEGY

A top-down strategic approach to EDM aligns business priorities to specific EDM components and capabilities. The approach should also evaluate current state of each relevant capability against the desired future state. An assessment based on a data management maturity model is a

good starting point for an EDM strategy and roadmap definition initiative. A data management maturity model assesses the above-mentioned capabilities with respect to the readiness of the firm from a people, policies, technology and adoption perspective. This helps the firm identify gaps and prioritize initiatives along a well designed road-map to achieve the future state.

6.1 BENEFITS OF DM PROGRAM IN BANKING

The benefits of anDM program to banking institutions are best analyzed from the lenses of the business capabilities that it supports and enables. Each business function potentially has its own unique needs around data and hence the benefits correspond to those unique perspectives. A few illustrative benefits are mentioned below.

■ **FOR OPERATIONS**, a centralized reference data management system will offer great advantages in providing accurate, timely and consistent data across systems. This will result in a huge reduction in reconciliation activities and will increase the efficiency and effectiveness of various teams.

■ **FOR RISK MANAGEMENT**, EDM offers among other things the ability to correctly identify counterparty risk. Accurate measurement and management of enterprise wide risk measurement and management would be virtually impossible without accurate, reliable and consistent data provided by an effective EDM.

■ **BENEFITS TO FINANCE AND ACCOUNTING** from EDM are obvious considering the performance analysis and management reports they produce that are viewed by external stakeholders (regulatory and market) and internal consumers (board, senior management and decision makers). EDM can allow these reports to be certified with a greater degree of confidence.

■ **DATA INTEGRITY AND CONSISTENCY**, which allow for greater confidence in the management reports and decisions, are of great importance from an audit, legal and compliance perspective as well.

■ **SALES AND MARKETING OPERATIONS** are immensely benefited from an EDM through the ability to have a single view of customer that enables effective cross-selling andup-selling.

7. CONCLUSION

The Effective Data Mining has become more important than ever before. A bank or financial institution embarking on an initiative to launch or revitalize its EDM program needs to keep in mind the following important aspects:

■ An efficient Effective Data Mining program should be designed to be in tune with the organization's own specific and unique business needs.

■ It is necessary to design a program that brings together stakeholders from both business and technology sides.

■ Technology solutions should be viewed as enablers of business capabilities and should be driven by business needs.

■ To develop, sustain and mature an Effective Data Mining program, a comprehensive framework including governance and control elements is needed.

■ It is important to maintain a balance between strategic long-term objectives and tactical quick wins.

■ A successful Effective Data Mining program is one which builds strong foundations, and at the same time allows for continuous evolution as business grows or transforms.

REFERENCES

[1]   Dr. Madan Lal Bhasin, 2006. Data Mining: A Competitive Tool in the Banking and Retail Industries

[2]   Berson, A., Smith, S., and Thearling, K. (1999). Building Data Mining Applications for CRM. McGraw-Hill, New York.

[3]   Ahmed, S. R. (2004). Applications of data mining in retail business. In In- formation Technology: Coding and Computing, International Conference on, volume 2, page 455, Los Alamitos, CA, USA. IEEE Computer Society.

[4]   Shaw, M. (2001). Knowledge management and data mining for marketing. Decision Support Systems, 31(1):127–137.

[5]   Giraud-Carrier, C. and Povel, O. (2003). Characterising data mining software. Intell. Data Anal., 7(3):181192.

[6]   Burez, J. and Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. Expert Systems with Applications, 36:4626–4636.

[7]   K. Chitra, B.Subashini, Customer Retention in Banking Sector using Predictive Data Mining Technique, International Conference on Information Technology, Alzaytoonah University, Amman, Jordan, www.zuj.edu.jo/conferences/icit11/paperlist/Papers/

[8]   K. Chitra, B.Subashini, Automatic Credit Approval using Classification Method, International Journal of Scientific & Engineering Research (IJSER), Volume 4, Issue 7, July-2013 2027 ISSN 2229-5518.

[9]   K. Chitra, B.Subashini, Fraud Detection in the Banking Sector, Proceedings of National Level Seminar on Globalization and its Emerging Trends, December 2012.

[10] K. Chitra, B.Subashini, An Efficient Algorithm for Detecting Credit Card Frauds, Proceedings of State Level Seminar on Emerging Trends in Banking Industry, March 2013.

[11] Lin, T. Y. (1994), "Anamoly Detection -- A Soft Computing Approach", Proceedings in the ACM SIGSAC New Security Paradigm Workshop, Aug 3-5, 1994,44-53. This paper reappeared in the Proceedings of 1994 National Computer Security Center Conference under the title "Fuzzy Patterns in data.

[12] Scott W. Ambler (2001) "Challenges with legacy data: Knowing your data enemy is the first step in overcoming it", Practice Leader, Agile Development, Rational Methods Group, IBM, 01 Jul 2001.

[13] Agrawal, R, and R. Srikant, "Privacy-preserving Data Mining," Proceedings of the ACM SIGMOD Conference, Dallas, TX, May 2000.

[14] Clifton, C., M. Kantarcioglu and J. Vaidya, "Defining Privacy for Data Mining," Purdue University, 2002 (see also Next Generation Data Mining Workshop, Baltimore, MD, November 2002).

[15] Gartner. Evolution of data mining, Gartner Group Advanced Technologies and Applications Research Note, 2/1/95.

[16] International Conferences on Knowledge Discovery in Databases and Data Mining (KDD'95-98), 1995-1998.

[17] R.J. Miller and Y. Yang. Association rules over interval data. SIGMOD'97, 452-461, Tucson, Arizona, 1997.

[18] Zaki, M.J., SPADE An Efficient Algorithm for Mining Frequent Sequences Machine Learning, 42(1) 31-60, 200