

Business Intelligence: Achieving Fineness through Data, Text and Web Mining

Jitendra Singh Tomar
Amity University
Sector-125, Gautam Buddha Nagar
Uttar Pradesh, India
Pin: 201313

ABSTRACT

There is no more a dearth of information with business organization incorporating ICT at core of its business structure. The amount of data available has been enormously increasing with business growth, hence determining patterns and trends out of substantial data is a challenge. The mining technologies are used by the organization to quest limitless data for crucial insight and knowledge. Web, Data, and Text mining are the important tools applied by the organization to automate finding hidden patterns to formulate policies and achieve competitive advantages in all functional business areas. Through mining techniques applied on various data repositories, the business intelligence systems along with analytical tools could present valuable and competitive information to the planners so as to develop new avenues for business growth. The usage of text, data, and web mining is discussed in this paper with a witness that how it can address business leadership and risk management, and enhance business intelligence.

General Terms

Your general terms must be any term which can be used for general classification of the submitted material such as Pattern Recognition, Security, Algorithms et. al.

Keywords

Data Mining, Text Mining, Web Mining, Business Intelligence, Information Systems, Knowledge Management.

1. INTRODUCTION

With the advancements in unified technologies, we have become a citizen of the world where information accessibility is at ease and the world is driven by the information. No longer we live in the world which is starving of information. Millions of data repositories are accessible by people and the organizations. In addition to the accessing the information, people and organizations are adding and contributing to the sea of information over the medium where there is no dearth of space. The web 2.0 along with ICT has made this possible. On the other side, need and analysis of information from these vast amounts of data and information is enormously tedious for the organizations in the information driven world. The organizations are always under pressure to thrive for new policy in ever changing business economy and to create these policies business world is relying on automated mining tools to mine web, data, and text, to meet the business requirements of all stakeholders.

Data Mining (DM) is defined as the practice of viewing information from new dimension or perspective, a perspective from which the information is not analyzed previously at any instance. It is a method of finding hidden patterns of information from a data repository to form new strategies and

policies. The prime objective is to establish a trend from historical data, establish a relationship among the data and the model that suits the dataset. DM involves varied fields of computer science, statistical modeling, artificial intelligence, information science, and machine learning [1]. The prime use of DM is business intelligence and risk management [2], which is crucial for critical decision making based on the relevant information in all business functional areas.

TM is related field to DM, but differs in its techniques and methodologies used. It deals with textual data rather than records. TM is defined as detection of hidden patterns, traits, or unknown information from textual data [3] that makes up huge quantity of data found on World Wide Web WWW. TM uses complex Natural Language Processing (NLP) techniques and incorporates computational linguistics, statistics, and machine learning. TM involves linguistically and semantically analysis of plain text, in realizing hidden traits in the text, like frequency of use of specific words, entity extractions, and documents summarizations.

The third mining technique is Web Mining (WM). WM is extraction of relevant information from the relics and mannerism recorded by the stakeholders on WWW. WM works along with the intelligent agents to certain targets, like competitors sites' [4]. These agents collect information from the host web server and collect as much information from analyzing the web page itself by visiting the hyperlinks, cookies, and recording traffic patterns. WM could be used for tracking customers' online behavior and the organizations could use the collected knowledge to establish better customer relationships, offers, and target potential buyers with exclusive deals. The dynamism of WWW could be dealt with ease through effective web crawling that could yield effective results.

In this paper, the prime mining technologies in information systems are discussed which are used by business organizations to achieve business leadership and risk management by exemplifying their experience using mining techniques. In addition, the paper discusses the major hindrances in effective application of data, web, and text mining today.

2. MINING AND KNOWLEDGE GO HAND IN HAND

Data Mining is a prime process of knowledge management and is considered as analysis step of Knowledge Discovery in Databases (KDD) process which comprise of three steps namely Pre-processing, mining, and data verification [5]. For mining process, the data stored in repositories are used for investigations and analysis supported by Artificial Intelligence and Neural Network predictive models. The AI techniques in

information systems are highly used in numerous business models for predictions and forecasting [6]. Also, the business risks and possibilities could be represented through decision trees which are rationale thinking tools [7] and can help for classifying information. One of the prominent tasks of DM is to classify the information based on the business domain in which it is applied. The analysis for classification is based on induced rules and relationship among the stored information and could lead to discovering unknown, unobserved and hidden patterns of the information possessed by an organization. This could help the organization in establishing foremost policies like locking in suppliers, establishing alliance, creating new product & services, raising barriers to market entrants, reducing cost of product and services, and many similar but prominent business strategies through establishing correlation. The companies can utilize the information systems for regression or predictive modeling in envisaging the trends and forecasts [8]. The regression in DM could be used in Exploratory Data Analysis by linking multi-dimensional data in tandem. Data visualization done through regression is the basis for the development of interactive DM tools in dynamic software which could summarize, extract, and aggregate information into factual figures [9].

DM entails huge computational resources like database or data warehouse which stores data to be mined and analyzed. For business activities, data is extracted, transformed and loaded through ETL processes by the organizations. Also, for these big computational activities, an IT infrastructure of substantial processing power is required. With changing times and growing concept & facility of cloud computing, organizations are combining the data mining services with cloud computing. DM is amalgamation of neuron networks with Artificial Intelligence, complex algorithms, parallel processing, and huge computational resources. There had been efforts to standardize DM and the Cross Industry Standard Process for Data Mining or CRISP-DM for short, was developed for the same in 1997 by the ESPRIT funding initiative under the consortium of SPSS (IBM) and NCR Corporation. The CRISP-DM standard is a leading prevalent standard for the development of DM software.

TM is a method to read and familiarize the text by the computers for retrieval of relevant documents, data, or text available on WWW using Information Retrieval (IR) Systems [10]. This is followed by NLP which is quite complex in nature and is the most difficult part of TM and Artificial Intelligence plays a vital role in NLP to perform the operations of TM with the same level of expertise as that of humans. The text is parsed and comprehended and also the grammatical construct of the text is understood. This is followed by the next step of extraction which is done using the linguistic tools to find out the correlation among various pieces of information that is comprehended and the hidden patterns in data are revealed.

The emulation of humans by TM techniques brews it as the most complex procedure. These techniques deals with highly unstructured data and hence its applications are also complex. In healthcare, it is used to link available historical medical records together to find a relation between symptoms and prescriptions [8]. TM process helps in clustering historical data into predefined semantic categories [11]. TM is effectively used in text summarizations by identification of grammatical construct of the text documents. TM is integral part of major IR search and smart engines which correlate various search queries to produce definite answer to the queries of the user. It is also used in TM OLAP as a textual

search tool rather than numerical. In WM operations, Internet and agent technologies are the basis behind WM that are based on fuzzy logic techniques. The IR tools that are supported by Semantic Web Technologies provide the course for the intelligent agents to explore and search WWW [12]. For easing out the retrieval process of information from WWW, the software are developed with semantics over web published content on the WWW and are continuously growing to enhance the searching process. The intelligent agents are programmed to perform mining techniques by analyzing the HTML document, parsing, and extracting all information from various hyperlinks, multimedia, and other content on the web. The tracking of online account and user preferences is done to identify the mannerism of the users to provide them with the most relevant business content. The sessions and transactions of the users are logged, data traffic along with various activities is logged and analyzed to give the stakeholder best and non-tedious experience to work on web as per their preferences and habits.

WM is divided into three main categories as given [13].

- Web content mining WCM – Locate and retrieve text, multimedia, and hyperlinks in the context of the search.
- Web structure mining WSM – Analyze the traffic flow and site maps of certain sites to explore the movement of the users.
- Web usage mining WUM - collects browser history, bookmarks, site logs, cookies, and metadata of the users. It may also be used to mine social network mannerism of users [14].

All three tasks are web oriented and are important for investigation of network, intranets, and VPNs. It is growingly used in multimedia mining to mine pictures, graphics, movies, audio, and its applications.

2.1 Mining Applications and Software

In 2008, the business intelligence software market reached over 7.8 billion USD [15]. One of the most widely used software for research & development is IBM SPSS and is one of the best business intelligence software, offering mining capabilities. Primitive to SPSS was Clementine, a graphical software that ruled the business intelligent market in 1990s. With cloud computing and knowledge process outsourcing developing fast, knowledge management companies have started providing business intelligence software on/off cloud to various other organizations. IBM provides online services and sophisticated tools for WM, called Surfaid Analytics [16]. They are also one of the prominent purveyors of solution-oriented packages such as IBM's Cognos 8 solutions [17]. Oracle too has business intelligence tools which are part of companies RDBMS software and are floated in the market as Oracle Data Mining. There are other enterprise solutions which are categorically business information systems like SAS Enterprise Miner offered by SAS [8], but the most prominent vendor is business information systems is SAP which offers ERP solutions that could be integrated with mining tools provided by them. Also, there are third party vendors like ATOS who use the platform of SAP software and work as information technology service provider [18]. Microsoft also offers platform dependant solutions by brand name of SQL Server Analysis Services. It has other important mining tools in its product line such as PowerPivot for SMEs. There are many open source mining tools available in the market like the Waikato Environment for Knowledge

Analysis (WEKA) [19], RapidMiner, and KNIME, but are not prioritized for usage as their usage is quite limited as compared to the open source software [20].

With the new enhancements in the technology, the cloud computing based mining tools for SMEs are utilized in current times [21]. The basic rationale behind preference of using cloud computing based mining tools is the fact that small and middle-sized enterprises lack the infrastructure and budget for installing the organization specific mining tools but cloud computing resolve this limitation. Cloud computing may amalgamate data warehouse and sophisticated mining and analysis tools as per the enterprise requirements, and that too with the expertise that may not be available with the small start-ups otherwise since start-ups and small enterprises may lack not only the financial resources but also the human resources and expertise in the IT field. The Infrastructure-As-Service IAS provides the middle-sized enterprises freedom of managing IT resources like software, hardware, and human resources management as well. On the downside, the high level of dependency and privacy issues are the major drawbacks of cloud computing as the information resource is stored and managed through the IT infrastructure of outsourced party. Due to dependency on IT infrastructure, the enterprise may get tied up with what the service provider is offering. Also, the enterprise's data is technically not under its control and possession. Also, the data could be misused and potentially could give rise to privacy and confidentiality concern.

3 BUSINESS INTELLIGENCE THROUGH DATA, TEXT, AND WEB MINING.

With rise of technology and availability of soft information, business organizations use mining software solutions in analytics and decision making process and hence business applications incorporate mining tools for enterprise decision making. Using mining techniques to acquire and analyze information is Business Intelligence (BI). With current business scenario, enterprise datasets are growing rapidly which is supported by information systems and effective DBMSs. As the amount of datasets mount up with growth of business, the traditional OLAP and manual analytical systems are becoming futile. The probable comfort is given by the BI applications and tools. "Data mining and artificial intelligence are at the top five key technology areas that will clearly have a major impact across a wide range of industries within the next three to five years" [22]. 80% of data found in enterprises information systems is unstructured, and would very likely to double in size nearly every three months [23]. BI application has modulated DSS for effective decision making. With the convenience and versatility, BI is now dominating all business functional areas and domains including retail, banking, and insurance industry [24]. BI software are widely used by knowledge workers, mainly executives, analysts, middle management, and operational management.

Mining tools reinforce Business Intelligence which leads to effective Strategic Information System (SIS) to gain competitive edge and effective risk management. The application of SIS could lead to strengthening customer engagements. Through BI application and mining of customer data, enterprises could know customer habits, patterns, and trends to practice effective CRM. This helps in segmenting the customers, knowing what they target and providing them with the customized product catalogue and the advertisements of the products which they may prefer. The mining of

customer data let the business enterprises identify the satisfaction levels of the customers and customer churning, if there is any. Currently the economies around the world today are information driven and are knowledge based economies [25] and mining tools reinforcing the BI applications is the main reason behind the development of information systems to support business enterprises.

It is essential for businesses to forecast and proactive in their policies. Through effective mining for CRM, enterprise can do the profiling of the customers based on their current and historical behaviors and their expected purchasing actions could be envisaged. Mining is used in predicting one of the most dynamic business environments of stock market. NASDAQ has hugely invested, approximately 450 USD, in BI applications [26] for managing its information systems. Their software solution incorporates smart BI agents for autonomous buying and selling of stock with minimal human intervention only in case there are erroneous results. The web page of Google Finance Service uses BI applications [27] to present graphs & charts to its end users, which is used for further decision making by the followers of Google. Keeping check over fraud detection, span monitoring, and likewise jobs is done using mining tools [28] and well know enterprises such as AT&T, bank of America, and likewise are using BI and mining techniques for fraud detection.

3.1 Achieving Competitive Advantages

In a cut throat competition, competitive pressure from adversaries and new entrants leads to competitive advantage [24]. The enterprises tend to build policies which will give the edge over others by analyzing large pool of information and acquire knowledge. The companies perform rigorous market research through mining and evaluate which products dominate the market and the obvious reasons behind the high sales. Media management companies do the market research to establish a correlation and commonality between program schedule slots and the type of audience prior to airing the program. They use automated mining tools to schedule the program slots and the result were better than the human manual scheduling [29]. Retailers are using mining tools to perform market basket analysis and introduce innovative baskets which are visible on the shelves of retail majors like WalMart, Costco, and K-Mart.

The financial institutions are using BI and mining for better risk management. Banks use the mining tools and BI applications for bankruptcy predictions [30]. Almost all major banks use the automated tools, based on neural networks, in their bankruptcy predictions by weighing the capability of the person and his liability to pay back the loan given. These BI applications help banking industry in reducing the risk of disbursing loans to the troubled customers. These applications also correlate the risk which banks can take and the assigned interest rates for lending. The tools also help the banks in assessing the credit loss with current loan portfolio of the bank. The head of bankruptcy was experience by banks during the financial crisis of 2008 and experts advocated for strict regulations to overcome such a scenario. The BI applications and tools in current times help the banks to create sophisticated profile of the customers before disbursing loans to them. Bankruptcy tools are also used by accounting companies [30] and are amalgamated with auditing tools that use BI to see the possibility of going bankrupt in future. Organizations like J.P. Morgan utilize their own bankruptcy tool called CreditMetrics [31], which monitors credit quality ratings and estimate possible risk incurred in the loan.

Automation of processes and Just-In-Time (JIT) are demand of current times in manufacturing establishments which is done through sophisticated Management Information Systems (MIS). The big reports generated by MIS tools at various levels of manufacturing goes as input to BI tools for better optimization of manufacturing processes [32]. The manufacturing performance could only be done through BI applications and are improvement over the MIS tools. The BI applications and tools could utilize the operation data from MIS tools to optimize the manufacturing process by mining the performance data. The reports of Material Requirements Planning and reports from inventory systems could produce huge amount of data which could be mined for enhancing the efficiency and proficiency of manufacturing organizations. The adaptive manufacturing schedules could be developed using BI tools to meet the requirements and fulfill the demands [32]. BI tools could be used to reduce the production wastage, unnecessary expenditures, overhead costs, optimize pricing analysis, examine sales, production, and adjustment of commodity price. These tools could also help the organization in doing warranty analysis; examine warranty claims and lawsuits occurring from manufacturing errors, faulty parts, and other possible risk factors. The manufacturing organizations could forecast possible risks and preclude flawed strategies and reduce errors through use of BI applications.

3.2 Improved CRM

With rise of technology and internet, the current generation of customer is well informed about the market and is well demanding too. Dealing with the customer in current times is getting more complex. The companies now have a well developed system to handle the customers in a better manner using the technology with a prime motive of customers' satisfaction and strengthening the relationship [33]. Companies are using CRM software to automate the customers' information processing, but with amalgamation of BI tools, the information could be viewed and analyzed from new dimensions. Customer satisfaction, strategy to attract new customers, customer churning, and the similar observations could be answered through better analysis of the information base using BI tools. BI tools are used by retailers, e-tailers, services providers, and telecommunications to analyze customers' service and target market and do effective product evaluation.

The CRM data is highly exploited since business organizations are utilizing BI tools profoundly to gain competitive advantage through customer satisfaction [34]. Enterprises are focused on providing value propositions like lower cost, offers, baskets, discounts, privileges, through rigorous data and information mining using BI tools which could prove them step ahead of their competitors. Internet has given the customer a self exploration facility through various search engines to search for the best deal available and hence for the companies to formulate the lower cost policy is difficult. The companies may well use BI tools to mine their information to offer the same set of products at various prices in different demography based on market sentiments. This price discrimination [34] allows the companies to place its product to appeal to its diverse customers in various demography. The products are placed in accordance to its competitors who are operating in the same area. The BI tools enable the enterprises to evaluate its customers as well as competition and place them safely in market. The BI tools let organization to mine the profile of customers it is targeting by evaluating their purchasing and spending mannerism over a

period of time as per their personal preferences. The company may fix the selling price of the product for specific category of customers to a point which they will surely pay. For rest of the customers, who will prefer the commodity but are willing to pay less, the company may float a limited period offer, thus try to engage with every possible and potential customer. Currently, companies use BI techniques to achieve effective price discrimination to fulfill the requirement of diversified and knowledgeable customers.

BI helps the companies in assessing the reasons behind the customer churning. Lower customer satisfaction due to various reasons is correlated to customer churning. It could be taken as the major indicator referring to organization's health. The reasons of customer dissatisfaction are identified through mining as these leads to customer switching to the competitors and ill proliferation of company's standing and brand. Churn analysis can also show reasons and causes for these turn over, like turning to rival companies, or finding substitutes to the enterprise products.

BI is also used in collaborative commodity selling through market basket analysis. Specific baskets could be constructed for certain group of users to suit their specific needs after analysis of CRM data. Market baskets offered by online retailers are more effectual where possibility of collaborative selling is higher by connecting diversified vendors together. This is bit difficult to do with physical market vendors due to their location and non-visibility to each other.

BI tools are also used in keeping track of the experience of the customers. While profiling the customers, their liking, satisfaction, and expectations is recorded through Customers Experience Management Systems which in fact is widely used by Service Industry. The CEM systems can help in identification of alpha customers who are highly valued so as to make them feel privileged and taking special care in managing them. Organizations are also adopting Web Mining (WM) to mine their customer's data and acquire valuable insights prior to creating any policy for the customers and business. The usage of RM is limited to WWW and sometimes is criticized since it possesses threats to individuals' privacy, anonymity, confidentiality, and security.

3.3 Improved Inventory Management and Logistics.

BI tools help the organizations to practice better control and management over their logistics and Supply Chain Management (SCM). Higher the efficiency of SCM better is the profitability of the company. BI tools amalgamate Warehouse Management Systems and Inventory Management Systems [35] to identify patterns, shortages, possible overproduction, and underproduction, demand spikes which are quite important reports for efficient logistics management. These parameters are important for distributors and customers satisfactions, on demand supplies, and quick response to inventory levels. Using BI tools, help supports supply chain management in managing these parameters. Current market situation could be mined using BI application for predicting the requirements of managing inventories, keeping under-flows and over-flows at par. Organization can have complete control over forecasting by analyzing market history, business events, rival vengeance, by using BI tools. Demand and supply cycles could be automated and appropriately managed for automatic store replenishment and effective warehouse management.

Logistics can affect inventories (under-flows or over-flows), specifically from slow response to spikes in demands and lack of precise forecasting. This is because the solutions to this process involve complex mathematical models which are tedious to solve. BI information systems automate these processes and the predictions are fast and precise. Using association rules and clustering, BI can help in effectively routing the supply effectively through the available channels [36]. Organizations use sophisticated BI tools to predict market segments demands that could be used in niche marketing.

The timely decision support in logistics management is very important and BI is helping the organizations in mining the information for precise predictions and timely decision support, especially in food supply chain [37] where the live stock could die at any of the various stages of supply chain and hence timely deliverance of food supply is very important. The slow reactions in food supply chain could give low yield and massive loss could incur to the organization. The Death-On-Arrival (DOA) could occur for known or unknown reasons at any level of supply chain but could be avoided if the trend analysis is done and information is mined about the possibility of food getting rotten during supply. BI tools can effectively identify such signs of dying livestock at an early stage and organization could plan appropriate action to overcome the problem and rescuing the remaining livestock. BI dynamically help in managing the supply of food products like milk, fruits, raw food material and help in planning and preservation of livestock management during the storage and distribution and help the organizations in time management of supply chain.

Currently supply chain business is majorly controlled by Third Party Logistics companies with whom various organizations deem contract for their supply of goods, materials, and commodities. The dedicated logistic companies use BI tools to provide effective services to the client organizations [38]. Through BI, these logistic companies are able to produce detailed reports and do effective forecasting. The companies are able to do the cost-benefit analysis while offering its services to the client companies. In addition, through BI tools, the visibility of supply chain is enabled that allows the users to track their supplies in the dynamic business environment in real time.

3.4 Fraud Detection and impeding Anomalies.

Possible patterns of fraud and anomaly in the daily or long term transactions could be detected using BI tools. These tools could mine the unusual and strange transactions and hence are a good help in forensic domain to detect fraudulence. Telecommunication and credit industry suffer the most fraudulence cases and are plagued by fraudulence and anomalies in their transactions [28]. BI tools are of great help to detect such possible cases. What separates fraudulence from anomalies is the operational behavior as anomalies are not deliberate actions like fraudulence. Anomalies may incur due to casual operations, corrupt data, glitches in operations, and technical errors and could prove fatal in information management. Anomalies should be audited so as to rectify the fracas and up heal the data integrity suffered due to shortfall. Fraudulent are deliberate actions of the white collar defaulters who are well versed in working and using the technology. The profile of such users and their actions on the technology for performing derogatory actions are identified by the BI applications to make organizations well aware of the possible

defaults that may occur if these people are partnering with these companies.

The credit companies use BI tools and application to record the behavior pattern of the customers. These tools are helpful in monitoring and mining erroneous behaviors and the transactions done by the users, especially in cases of stolen credit cards. The transactions post losing the credit card are done within few hours which also see the numerous transactions for large amount of money. The pattern of such uses could be mined prior to releasing the credit card to the customers as BI fraud detection can easily spot such anomalies. BI tools can detect such suspected behaviors over a length of time, pointing which accounts and individual or customers require special attention. In financial crimes too, BI applications are providing great help. The patterns about money laundering, sifting records of individuals, insider trading, intrusion detection, spam detection may be recorded by BI system to caution the organization prior to any specific commitment. With all the definite advantages which the BI tools have, improper predications by BI tools may lead to legal consequences, if not handled appropriately.

4 MINING TECHNOLOGIES AND CHALLENGES.

BI tools are specialized application which works with information systems to strengthen company in crucial mining of information for better strategy implementation and planning. The development of these applications involves high cost and requires specialized human resource to work on them. Currently industry faces dearth in availability of trained manpower to work on these applications as most business managers specialize in manual techniques [39].

Even the availability of appropriate BI application with the entire needful infrastructure poses certain challenges to the organizations. The challenges are broadly classified into three categories:

- Technical Challenges,
- Ethical Challenges, and
- Legal Challenges.

Technical Challenges: Specialized infrastructure is required for implementation of BI application, supported by people of specific competency. Foremost requirement for mining is to have a data warehouse for the structured storage of information from where it could be accessed by various managerial levels of the organizations for appropriate decision making. Technical infrastructure needed for ease of mining include high end servers, distributed architecture, network software and hardware tools, AI packages including NLP, IR engines, and the list is limitless. The configuration too is quite versatile based on affordability. Also, majority of the BI applications are area specific and are non extendable and use specific models and techniques that explicitly could be used in specific areas such as data visualization, market analysis, fault detection, and may not be suiting other areas. A lack of unified platform of BI tools to satisfy business needs at large is missing. Also, the working interface of BI software is bit complex and tedious to work upon and hence the executive may avoid working on the software. Scalability is also one major limitation with these tools. With the growth of business these software may not be able to scale with the business and support it.

Ethical Challenges: For mining, organizations are using the data of individuals which may include private data, and hence raise concerns around how such data should be used. The mining tools identify the history and mannerism of potential customers and breaching privacy and confidentially. With the awareness rising among the global citizens who use the soft platforms like internet, these uses of mining information through mining tools made public aware of their legal and privacy rights [40]. Countries like US are passing legal regulations like Health Insurance Portability and Accountability Act (HIPAA) where a prior consent from the individuals has to be taken for usage of their information.

Legal Challenges: Use of information technology has enhanced the data quality by structuring data as authentic, complete, accurate, and data is legally liable for its authenticity [41]. It is believed that more than 25% of critical data in top companies' databases are inaccurate and incomplete [42]. The efficiency and accuracy of predictions done by BI is based on quality information input which it will mine and in case the input information to these BI tools is inappropriate, the end result may not be up to the mark. The quality of BI reports is as good as the data quality of the input, better and accurate data yields better BI decisions. In case the companies are not able to generate appropriate output report through BI applications as data supplied to BI applications is incorrect, companies may face legal action especially in case if the output reports deals with tax figures or if the reports are fiscal report. The respective governments of the countries have clear mandate on such reports and if the company willingly or unwillingly is not able to produce actual figures, it may face the legal action under legislation of the respective country.

5 CONCLUSION

The rise of technology and business dependency on it has obligate the business organization to apply efficient tools for forecasting and BI tools gives the companies a breathing space. BI is driving the competitiveness in the current business scenario through appropriate mining of information. The sea of information is available to the organization, both internally as well as externally, on their intranet and on internet. The companies are bound to explore the available information to formulate specific strategy to give itself a new dimension.

Enterprises utilize the vast amount of data to achieve competitive advantage, better CRM, effective ERP, efficient resource planning, and fraudulence detection through BI tools and applications. The BI tools are considered as specialized applications that incorporates AI techniques, decision trees, NLP, and SM technologies. With all the benefits which BI tools can bring to a business, many challenges impede the advancements of such tools which include technical challenges, ethical challenges, and legal challenges. More enterprises are including BI applications in their profiles, more progress, novelty, and innovation is involved in developing BI applications to suit overall requirements of an organization. The enterprises will be highly dependent of BI tools which may be amalgamated with the information systems of future to help business in all round operations.

6 REFERENCES

[1] Bill Palace, (1996) "Technology Note prepared for Management 274A" Anderson Graduate School of Management at UCLA.

- [2] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, (2008) "The Elements of Statistical Learning: Data Mining, Inference and Prediction," New York, Springer-Verlag, ISBN- 0 387 95284-5
- [3] Marti Hearst, (2003) "What Is Text Mining?" SIMS, UC Berkeley.
- [4] Prof. Anita Wasilewska, (2011) "Web Mining Presentation 1" CSE 590 Data Mining, Stony Brook.
- [5] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, (1996) "From Data Mining to Knowledge Discovery in Databases," AI Magazine, American Association for Artificial Intelligence AAAI, Vol. 17 No. 3.
- [6] Meryem Duygun Fethi, Fotios Pasiouras (2010) "Assessing Bank Efficiency and Performance with Operational Research and Artificial Intelligence Techniques: A survey" European Journal of Operational Research, pp. 189–198, Elsevier.
- [7] J.R. Quinlan, (1986) "Induction of Decision Trees", Machine Learning, Kluwer Academic Publishers, Boston.
- [8] Dave Smith (2010) "Using Data and Text Mining to Drive Innovation" PhUSE 2010, UK.
- [9] Michael Goebel, Le Gruenwald, (1999) "A Survey Of Data Mining and Knowledge Discovery Software Tools," SIGKDD Explorations, Vol. 1, Issue1. Pg 20, ACM SIGKDD.
- [10] Judy Redfearn and the JISC Communications team, (2006) "What Text Mining can do" Briefing paper, 'Joint Information Systems Committee' JISC.
- [11] Neto, J., Santos, A., Kaestner, C., Freitas, A. 2000) "Document Clustering and Text Summarization" In the Proceeding of the 4th International Conference Practical Applications of Knowledge Discovery and Data Mining PADD-2000, London, UK.
- [12] Rafael Berlanga, Oscar Romero, Alkis Simitsis, Victoria Nebot, Torben Bach Pedersen, Alberto Abelló, María José Aramburu (2012) "Semantic Web Technologies for Business Intelligence" IGI.
- [13] Sankar K. Pal, Varun Talwar, Pabitra Mitra, (2002) "Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions" IEEE Transactions on Neural Networks, Vol. 13, No. 5.
- [14] Federico Michele Facca, Pier Luca Lanzi, (2005) "Mining interesting knowledge from weblogs: a survey" Data & Knowledge Engineering, 53, Elsevier.
- [15] Ralf Mikut, and Markus Reischl, (2011) "Data mining tools" Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 1, Issue 5.
- [16] IBM, SurfAid Analytics (2003).
- [17] IBM Software Group Case Study (2010) "Great Canadian Gaming Corporation Leverages IBM Cognos 8: Solutions for Financial Consolidation and Reporting Standardization".
- [18] Atos, (2011) "Business Intelligence solutions: Decisions that are Better-Informed Leading to Long Term Competitive Advantage".

- [19] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten (2009) "The WEKA data mining software: an update" SIGKDD Explorer News.
- [20] Christian Thomsen, Torben Bach Pedersen (2009) "A Survey of Open Source Tools for Business Intelligence" International Journal of Data Warehousing and Mining, Vol. 5, Issue 3, IGI Global.
- [21] Oksana Grabova, Jerome Darmont, Jean-Hugues Chauchat, Iryna Zolotaryova (2010) "Business Intelligence for Small and Middle-Sized Enterprises" SIGMOD Rec. 39.
- [22] Sang Jun Lee, Keng Siau, (2001) "A Review of Data Mining Techniques" Industrial Management and Data Systems, 101/1, MCB University Press.
- [23] Shantanu Godbole, Shourya Roy, (2008) "Text Classification, Business Intelligence, and Interactivity: Automating C-Sat Analysis for Services Industry" KDD'08, ACM Las Vegas, USA.
- [24] Thiagarajan Ramakrishnan, Mary C. Jones, Anna Sidorova, (2011) "Factors Influencing Business Intelligence and Data Collection Strategies: An empirical investigation", Decision Support Systems.
- [25] Aura-Mihaela Mocanu, Daniela Litan, Stefan Olaru, A. Munteanu (2010) "Information Systems in the Knowledge Based Economy" WSEAS Transactions on Business and Economics, Issue 1, Vol. 7.
- [26] K. Laundon and J. Laundon (2012) "Enhancing Decision Making" Managing Information Systems: Managing the Digital Firm, Pearson Education, Pearson Hall.
- [27] Byung-Kwon Park and Il-Yeol Song (2011) "Toward total business intelligence incorporating structured and unstructured data" In Proceedings of the 2nd International Workshop on Business intelligence and the WEB (BEWEB '11), ACM, NY, USA.
- [28] Clifton Phua, Vincent Lee, Kate Smith, Ross Gayler, (2010) "A Comprehensive Survey of Data Mining-based Fraud Detection Research" Cornell University library, CoRR.
- [29] Fitzsimons, M., Khabaza, T., and Shearer, C. (1993) "The Application of Rule Induction and Neural Networks for Television Audience Prediction" In Proceedings of ESOMAR/EMAC/AFM Symposium on Information Based Decision Making in Marketing, Paris, pp 69-82.
- [30] Amir F. Atiya, (2001) "Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results" IEEE Transactions on Neural Networks, vol. 12, no. 4.
- [31] M. Crouhy, D. Galai, and R. Mark, (2000) "A comparative analysis of current credit risk models," J. Banking & Finance, vol. 24, pp. 59–117.
- [32] MAIA Intelligence (2009) "Business Intelligence in Manufacturing".
- [33] Injazz J. Chen, K. Popovich, (2003) "Understanding Customer Relationship Management (CRM): People, process and technology", Business Process Management Journal, Vol. 9, pp.672–688.
- [34] Dien D. Phan, Douglas R. Vogel, (2010) "A Model of Customer Relationship Management and Business Intelligence Systems for Catalogue and Online Retailers", Information & Management, Vol. 47, Issue 2, Pages 69-77.
- [35] Mu-Chen Chen, Cheng-Lung Huang, Kai-Ying Chen, Hsiao-Pin Wu, (2005) "Aggregation of Orders in Distribution Centers using Data Mining" Expert Systems with Applications, Volume 28, Issue 3, Pages 453-460, Elsevier.
- [36] Van den Berg, J. P. (1999) "A literature survey on planning and control of warehousing systems" IIE Transactions, 31, PP.751–762.
- [37] Y. Li, M.R. Kramer, A.J.M. Beulens, J.G.A.J. Van Der Vorst (2010) "A Framework for Early Warning and Proactive Control Systems in Food Supply Chain Networks" Computers in Industry, Vol. 61, Issue 9, pp. 852-862.
- [38] Srinivasa Rao P, Saurabh Swarup (2001) "Business Intelligence and Logistics" Wipro Technologies.
- [39] Gregory Piatetsky-Shapiro, Ron Brachman, Tom Khabaza (1996) "An Overview of Issues in Developing Industrial Data Mining and Knowledge Discovery Applications" KDD-96 Proceedings.
- [40] K.A. Taipale (2003) "Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data" Columbia Science and Technology Law Review 5.
- [41] Carlos Rodríguez, Florian Daniel, F. Casati, Cinzia Cappiello (2010) "Toward Uncertain Business Intelligence: The Case of Key Indicators" Internet Computing, IEEE, vol.14, no.4, pp.32-40.
- [42] K. Laundon and J. Laundon (2011) "Foundations of Business Intelligence: Databases and Information Management" Managing Information Systems: Managing the Digital Firm, Pearson Education Inc.