

The Use Of Data Mining And Text Mining Techniques In Business Analytics

Ambreena Muneer¹, Aasifa Arabi², Qurat Ashraf³, Suhail Javid⁴,
Yasir Saleem Khan⁵

^{1, 2, 3, 4, 5} Dept of Computer Science Engineering

^{1, 2, 3, 4, 5} Iqbal Institute of Technology and Management

Abstract- *In this paper, the data mining and text mining techniques are used in business analytics and the way they support the use of business intelligence in e-businesses are discussed. The comparison between data mining, business intelligence and business analytics is also discussed. Suggestions on how e-businesses can influence on data mining to enhance their competitiveness are discussed.*

Keywords- Text Mining, Data Mining, Business Intelligence, e-Businesses,

I. INTRODUCTION

The information present in real time and the ability to make it readily accessible to everyone has been a major catalyst for many organizations to embrace globalization and improve productivity. For many consumers, they have benefited from the changing landscape that has evolved over the years in the business environment, in particular the new business models that have been introduced by many traditional companies as well as new the business world we work and operate in has changed dramatically over the past 20 years. With the computer and the Internet becoming ubiquitous tools in almost all business organizations now, the often-mentioned information age has really come of age. The capacity to collate data and companies. Consumers are now able to buy most products on the Internet.

We use the hand-phones to conduct meetings, order tickets, and check stock prices, among a host of other tasks. Computers, information systems and telecommunications have been the vanguard for these new business models. But often, not many of us realize that data is actually the primary driver for the information age. Without relevant data, the need for computers and communications would be much reduced. Companies are requiring more regular in-depth analytical review of the data within their repositories to have a better understanding of their business environment and competitiveness capability. The ability to generate valuable insights based on business data is termed business intelligence. The approaches to generate these insights are commonly the function of business analytics. Business analytics is the use of analysis techniques and decision rules to

provide business users with critical insights of the operational and performance characteristics in every aspects of the business. In this paper, the use of data mining techniques in business analytics and how they support the use of business intelligence in e-businesses are discussed. The distinction between data mining, business analytics and business intelligence is presented. Suggestions on how electronic businesses can leverage on data mining to enhance their competitiveness are discussed.

II. CONDUCTING BUSINESS IN THE INFORMATION AGE

The Internet, the increasing sophistication of customer needs, and relentless globalization are forcing businesses to change the way they manage business and customers. When you walk into supermarkets, have you ever wondered why some of them place eggs next to bread? When you receive product brochures from your credit card issuers, have you ever stopped to ponder why they are able to know what you desire? When you apply for car insurance, do you know why your insurance company requires all sorts of personal information about you and your car before deciding on your premium? In today's global competition, a good organization will make the effort to find out whether its customers are satisfied with its products and services. However, an excellent organization will not only know whether it has satisfied customers, but will also be able to understand why they are satisfied and what will make them stay satisfied in the long term. In this scenario, the former is focusing on symptoms while the latter is going beyond the symptoms and trying to understand the causes. The difference between these two organizations in having a more in-depth understanding of the customer lies in the amount of business intelligence one has over the other.

Furthermore, businesses are relying on the Internet to provide some if not most of their services other than the actual manufacturing of their products. Online companies like Amazon and E-Bay obviously based their business models solely on the Internet. However, traditional companies who are either selling tangible products or intangible services do also feel the need to have a presence on the Internet. The form

of presence could vary from informational, to marketing and selling, to after-sales services and customer support.

For all companies, a major portion of their revenues is generated from repeat customers. This is especially true for e-businesses. E-business faces intense competition on the Internet as customers are spoiled for choice and prices when doing online shopping. It is so much easier for a customer to browse and move from one virtual online store to another in a matter of clicks compare to walking and driving from one store front to another. Because of the nature of online shopping, building customer loyalty and understanding and anticipating customer needs become a lot more critical than in traditional brick-and-mortar organizations.

Data generated in online customer transactions, be it sales, queries, feedback, searches, or simply browsing provide the organizations with a potential wealth of information that could help reinforce loyalty for existing customers or create opportunities for new leads. Data generated in internal databases maintained by organizations could also be a source of relevant information to help organizations in improving business operations and enhance capabilities and competencies.

Thus, in the era of the information age, businesses have to view its collected data as a source of competitive advantage. Data mining and text mining are promising techniques to harness the potential value of data found in organizations. The application of data mining and other data analysis tools produces useful information or relational functions that helps managers to make insightful decisions.

III. DISTINCTION BETWEEN DATA MINING, BUSINESS ANALYTICS, BUSINESS INTELLIGENCE

Some sources use the terms data mining, business analytics and business intelligence in an interchangeable manner. However, in strict theoretical sense, there are distinct differences among the three terms.

Data mining is a collective term used to describe different analysis techniques such as statistics, artificial intelligence and machine learning that are employed to scan huge amounts of data found in the organization's databases or online databases. Its purpose is to identify patterns in the data set. These data sets could come from a single database or could come from integrated data established in a data warehouse. The key here is that one is searching for a pattern or relationship among different data groups. It is unlike a normal OLAP query whereby an identified pattern or relationship is used to process answers from the database. In

data mining, by identifying possible patterns, it could then help to predict, associate and group events, products, or customers in a more effective manner so that the organizations could provide better products or services to the customers or improve their operations.

Business analytics is used to describe the entire function of applying skills, technologies, standard practices, and algorithms related to data mining and data collation methods to generate valuable information, usually presented in highly readable format so that managers can make business decisions and to control and manage their business operations. Data mining is usually applied in the back-end of the business analytics function while the front-end of business analytics function consists of executive reporting metrics and collated information. When the business analytics function is efficiently and effectively executed, it may become a core competence for the organization in the form of valuable business intelligence that will support the strategic actions undertaken by the organization.

IV. USE OF DATA MINING AND TEXT MINING IN BUSINESS ANALYTICS

The main locomotive driving the application of business analytics in businesses is data mining or knowledge discovery in databases. In applying data mining, the business analytics professional act like a data detective, analyzing data to better understand an organization's current and past situation, predict future outcomes and act effectively. Customers' current and past transaction behaviors in sales, servicing and selection choices can also be scrutinized and grouped either in clusters. Event sequencing and product portfolio determination can also be studied using data mining.

Akin to flying an airplane, the use of business analytics is like having a management navigation dashboard to steer the business towards the right path and achieve greater performance.

Data mining uses algorithms to mine huge data sets so as to identify patterns in the data set that may be used to establish valuable information that help organizations to compete more effectively in the market place. By uncovering patterns in the data set, data mining can predict, associate and cluster events, products, or customers in a more effective manner so that the organization could provide better products or services to the customers or improve the efficiency of their operations.

To describe the power of data mining, let's use a simple analogy. In today's world, we have powerful database

software that can help us to do a variety of data queries. These data queries allow us to extract information from huge databases. But in order to perform the data queries, we must first know and state the query to be answered before the database software can help us to perform the specific query.

In data mining and text mining, it is the opposite. We do not know what queries we want to ask in advance. We simply use data mining to help us search through the huge databases in the organization and identify patterns in the data sets that might be useful to us. In simple analogy, database query helps us to obtain information that we know existed in the databases while data mining helps us to discover knowledge that we do not know existed in the databases. The latter is termed as knowledge discovery [1], a non-trivial process to identify valid, novel, practical and recognizable patterns in data residing in huge databases. The terms knowledge discovery and data mining are sometimes used interchangeably.

V. THE MECHANICS OF DATA MINING

The strategy for performing effective data mining generally encompasses four phases. In today's information intensive business environment, the availability of data is not an issue in most organizations. However, the formats in which these data are represented may not all be suitable for data mining purpose. Thus, the first phase of data mining strategy is to prepare the data. Data preparation is a critical activity and is often the most time consuming activity in data mining. It involves converting unsuitable data formats into specific formats that lend themselves readily for data mining. Data preparation may include tasks such as data cleaning (how to deal with missing data), data transformations (converting data values through normalization, mapping, and/or aggregation), and data reduction (combining data that involves large number of variables into a smaller set of variables).

Once data preparation is completed, the second phase involves selecting appropriate data mining and text mining techniques to search for patterns in the data sets. Depending on the objective of the data mining exercise, one would most likely use any one of the primary data mining techniques such as association, clustering and classification or estimation/prediction. Each approach invokes a particular algorithm that will systematically search for specific forms of pattern in the data sets.

In the third phase, the results generated by data mining have to be interpreted and the model it created has to be assessed for accuracy, validity and/or relevance. If the model proved to be accurate, valid and/or relevant, it will then be used as a decision making tool in business applications. If the assessment of the model concluded that it is insignificant,

the data mining exercise can be repeated using a bigger sample data set or alternatively using new data attributes.

In the fourth phase, once the association, clustering or classification model has been assessed and is deemed to be accurate, valid and/or relevant, deployment of business plans guided by the model will follow. For example, if a grocery chain develops an association model that shows there is a high percentage of people who purchase baby diapers on weekends also purchase sports magazines, the chain could deploy a business plan that make sports magazines an obvious display item next to diapers. This is a case of knowledge discovery followed by relevant

VI. LEVERAGING ON BUSINESS ANALYTICS AND BUSINESS INTELLIGENCE IN E-BUSINESS

Organizations are beginning to realize that the ability to generate business intelligence provides them with a competitive edge over their competitors. This is especially so for e-businesses. This phenomenon is primarily driven by the abundance of data created by advances in information and communication technologies as well as the increasing use of the Internet to conduct business operations. Various data mining techniques such as association, clustering and classification or estimation/prediction are currently deployed across many industries [2].

The data mining technique related to association is commonly applied in the retail industry. The main application here is to associate a basket of products that are most likely to be purchased by customers. This application is commonly called market basket analysis. Results from market basket analysis help retailers to cross sell products, develop focused promotions and design more effective shelf arrangement of products. A good example is the online book or music retailers whereby the moment you purchase or select a particular item, the online retailer will suggest various other items of similar genre to you. On websites, e-businesses can also place their products that have high association on the same webpage to entice online customers to increase their purchase volume. Association rules are usually derived using apriori algorithm [3] that detects important relationships among cross-tabulation tables.

Clustering, as a data mining technique, is commonly used by businesses to perform customer segmentation. Customer segmentation is usually used to support marketing promotions and target selling. However, it can also be used to identify customer segments that are good credit risks and those that are bad credit risks, especially in the insurance and banking industries. Such an application of clustering is termed

risk management. Clustering can also be applied to detect fraud. Such applications could be found in the credit card industry whereby data mining models are built to detect possible fraudulent credit card transactions. One of the most common techniques used in clustering is the K-Means algorithm [4], using Euclidean distance to identify distinct clusters in data.

For e-businesses, the ability to have real-time identification of customer segments, credit risks, or fraudulent transactions is critical for business competitiveness and to mitigate business risks. Figure 2 provides an illustration of how data mining can be deployed to detect fraudulent credit card transactions in e-businesses. In an online purchase scenario, credit card details will be entered in the website. The information will then be transmitted via Internet to a processing center for verification, authentication and approval. It is at the processing center whereby fraud detection models derived from data mining (clustering) can be applied to evaluate the possibility of fraudulent cards. With a quick assessment by the model, the card payment is either placed in the low probability or high probability categories and automated decision can be initiated as to whether final approval will be given to accept or reject card payment. Besides clustering and association, the other commonly used data mining approach in business applications is classification, estimation or prediction. Classification, estimation or prediction is based on induction-based supervised learning. In classification, the goal is to place records into defined classes. The dependent variable is normally categorical. In estimation, the goal is similar to classification. However, the dependent variable is normally numerical. Both methods attempt to predict in which defined class should a new instance be placed (an instance is a record in a data set). In prediction, the goal is to determine a future outcome rather than to place new instances in pre-defined classes. Using credit card as an analogy, to determine if a credit card purchase is fraudulent will be classification, to estimate the probability that a credit card has been stolen is estimation, and to forecast the number of credit cards that are likely to be stolen in the next 6 months is prediction. Although classification, estimation or prediction can be used in many applications, one of the most common applications is to build churn models.

In industry, churn modeling evaluates which customers are likely to leave in the near future. Decision tree [5], statistical regression and neural network are the common techniques employed in this data mining approach. Churn models are important in e-businesses. It allows management to determine the likelihood of an existing online customer from switching to a competitive firm for his/her purchase in the near future. If the likelihood is high in the near future,

proactive action could be initiated to prevent customer from switching to competitors. Such proactive actions include offering discounts, lower pricing, and other enticing propositions to the customer.

VII. CONCLUSION

Increasing data resources are expected to drive a growth in business analytics and thus data mining. Businesses are beginning to realize that the application of data mining and text mining provides them with a competitive edge. With data increasing in an exponential manner, the ability to use data mining to sieve through massive amount of data and identify relevant patterns will become a strategic tool in improving key areas of the business such as customers, operations and the supply chain.

REFERENCES

- [1] Agrawal, R., Imielinski, T., and Swami, A. (1993) Mining Association Rules Between Sets of Items in Large Databases. Proceedings of the ACM Sigmoid International Conference on Management of Data, New York.
- [2] Koh, H.C. (2005) Data Mining Applications for Small and Medium Enterprises, Centre for Research on Small Enterprise Development, Singapore.
- [3] Lloyd, S. P. (1982) Least Squares Quantization in PCM. IEEE Transactions on Information Theory, 28, 2. pp. 129-137.
- [4] Quinlan, J. R. (1986) Induction of Decision Trees. Machine Learning, 1, 1, pp. 81-106.
- [5] Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy, Editors (1996). Advances in Knowledge Discovery and Data Mining, The MIT Press.