



# Journal of Internet Banking and Commerce

*An open access Internet journal (<http://www.arraydev.com/commerce/jibc/>)*

*Journal of Internet Banking and Commerce, August 2014, vol. 19, no.2  
(<http://www.arraydev.com/commerce/jibc/>)*

## Modeling and Technical Analysis of Electronics Commerce and Predictive Analytics

---

**KAMAL NAIN CHOPRA, PhD**

**Professor in Applied Physics Department, Maharaja Agrasen Institute of Technology, GGSIP University, Rohini, New Delhi, India.**

*Postal Address:* Applied Physics Department, Maharaja Agrasen Institute of Technology, Rohini, GGSIP University, New Delhi - 110086, India

*Author's Personal/Organizational Website:* [www.mait.ac.in](http://www.mait.ac.in)

*Email:* [kchopra2003@gmail.com](mailto:kchopra2003@gmail.com)

Prof. Kamal Nain Chopra is the Professor in Applied Physics Department, Maharaja Agrasen Institute of Technology, GGSIP University, Rohini, New Delhi, India. After doing B.Sc from Delhi University, he has done M.Sc (Physics), M. Tech. (Opto-electronics), and Ph. D. (Applied Physics), all from the Indian Institute of Technology (IIT), Hauz Khas, Delhi, New Delhi, India. He has served Defence Research and Development Organisation (DRDO) in various capacities for a number of years. He has taught in Hans Raj College, Delhi University, and Netaji Subhas Institute of Technology (NSIT), Delhi University, besides serving as Research Scientist in various Projects of national importance at IIT, Delhi. He has more than 150 Research Papers in the International Journals of Science and Engineering. His areas of interest are Electronics Commerce, Benefits of BI, Factors influencing BI, Technology requirements, Designing and implementing BI, OLAP, Data Mining, representation technologies, Predictive Analytics, Decision Support System, and the related technologies, terminologies, and tools in Knowledge Driven Industries, besides Physics and Engineering Topics of modern interest.

---

### Abstract

Recently, the Electronics Commerce and Predictive Analytics has become the subject of much interest and research activity. The present communication brings out the technical analysis of the various intelligence techniques, and also the characterization of various parameters like - Web analytics, and the related technologies, terminologies, and tools.

The concepts of business intelligence like - benefits of Business Intelligence (BI), factors influencing BI, technology requirements, designing and implementing business intelligence, and the related parameters like - the data warehouse, online analytical processing (OLAP), Data Mining, representation technologies, and their role in improving the enterprise operation effectiveness have been discussed from the point of view of Information Technology. Various business research experts are pursuing the work on the modeling of some of these techniques rigorously. The predictive accuracy of the Predictive Modeling Methods has been briefly discussed. It is expected that the techniques described in the paper, and the technical discussions on the subject will be very useful to the new entrants in the field.

Keywords: Electronics Commerce, Benefits of BI, Factors influencing BI, Technology requirements, Designing and implementing BI, data warehouse, OLAP, Data Mining, representation technologies Predictive Analytics, Decision Support System ,Web analytics, and the related technologies, terminologies, and tools.

© Kamal Nain Chopra, 2014

---

## **INTRODUCTION**

It is widely known and accepted that in spite of having a lot of data, many business organizations do not know exactly how to use it for deriving and extracting the optimum value from this. Business Analytics is a science of techniques of modeling, analyzing and interpreting the data, by using the quantitative and computational approaches for finding the solutions to very complex problems, leading to the correct decision making. It has to be understood that there are various applications in the management of organizations including (i) exploring the novel ways for tackling the problems faced by managers; (ii) helping in the clarification of the objectives and especially priorities; (iii) suggesting and defining the alternative strategies for the optimisation of the profits and the other benefits. Analytics is a powerful tool for determining the future for business. This involves the various quantitative techniques for business decision making, which are becoming very important for better returns in the modern rapidly changing markets, constantly facing increasing competition, and also being challenged by the tighter margins. These techniques are based on the increased computing power to study the available data, and achieve accurate results, optimum performance, and high degree of success. These techniques are very useful for the Analytics experts having the ability of the creative thinking, and good communicate skill, as they have to work in close contact with their business clients from the small and big corporate houses, with different requirements and strategies. The principles and practices of Analytics are used by various organisations including banks and finance houses, transport systems, e.g. Railways, Airports, Airlines, Government bodies, big Engineering and Manufacturing industries, and most importantly by the various Management Consultancy companies

### **Business Analytics**

Business Analytics, also called Management Science or Operational Research, is an important part of any big organisation, which uses it for the major decisions, by taking

into consideration the large resource implications. Web analytics is mainly the science for studying the behavior of website visitors, though in the commercial context, it is employed for the use of the data collected from a web site, to find the aspects of the website for achieving the business objectives in both; the short term basis, and the long term basis. For the Web Analytics, we have to know the terminology, technology, tools, and even the case studies. The relevant technology is mainly based on Internet and TCP/IP, Client and Server computing, Hyper Text Transfer Protocol (HTTP), Server Log Files and Cookies, and Web Bugs. The Internet is the infrastructure which is instrumental in providing for the delivery of data between computer based processes. A cookie refers to the information that a Web site puts on our hard disk, so that it can remember something about us at a later time. In other words, it is the information for future use, which is stored by the *server* on the client side of a client/server communication, and hence a cookie records our preferences when using a particular site. Cookies are used to solve the Statelessness of the HTTP Protocol, and also to store and retrieve the user specific information on the web. The technique is quite simple: When the HTTP server responds to a certain request, it sends the additional information that has been stored by the client - state information. Also, when the client sends a request to this server, the client returns the cookie that contains its state information, which may be a client ID, used as an index to a client data record on the server. The Server Log Files consist of the Passive data collection, and the Normal part of web browser/web server transaction. The Page Tagging has two important components - Active data collection, and the third party to implement - a vendor. Based on the (HTTP) system, each request for a Web page is independent of all other requests, and that is why, the Web page server has no memory of what pages, it has sent to a user previously or anything about our previous visits. Thus, the cookie can be understood as a mechanism, which allows the server to store its own information about a user on the user's computer. TCP/IP are the protocols, which provide for the reliable delivery of data on the Internet.

Other considerations are: (i) Client's sending a request to a server; and (ii) Server's sending a response to the client. In the beginning, the connectionless client opens the connection to server, and sends the request, and subsequently, the server responds to the request, and finally closes the connection. The important characteristics of the system are that the Stateless Client and the Server have no memory of the prior connections, and is just unable to distinguish one client request from another client. Server Log Files also are very important in this analysis.

The technical issues for the server log data are: (i) Data Preparation, (ii) Page view Identification, (iii) User Identification, and (iv) Session Identification. The data collection uses the Click stream (Click stream is a sequential series of page view requests made from a single user.) by studying the Server Log Files and Page Tagging. The main components of Data analysis are: Data Preparation, Pattern Discovery, and Pattern Analysis, each of which has to be handled intelligently for reaching the optimum decision.

The important components of the technique are: (i) Web Site, which shows the collection of interlinked Web pages, including a host page, residing at the same network location.; (ii) User or Visitors , who use a client to interactively retrieve and render resources or resource manifestations, by accessing files from a Web server, with the help of a

browser; (iii) User Session , which refers to a delimited set of user clicks across one or more Web servers; (iv) Server Session or Visit, which involves a collection of user clicks to a single Web server during a user session; and (v) Page view , which provides the visual rendering of a Web page in a specific environment at a specific point in time. It is important to note that a page view consists of several items - frames, text, graphics, and scripts which construct a single Web page. Some other important parameters for the technical analysis are: (i) Conversion Rate, which gives the number of completers divided by the number of starters for any online activity that is more than one logical step in length, like Purchase or Downloading a research article; (ii) Abandonment Rate, which for any step in a multi step process is one minus the number of units that lead to taking it to step (n+1) divided by those at step n, the formula being given by:

$$\text{Abandonment Rate} = [1 - \{(n+1)/n\}] \quad \text{----} \quad (1);$$

(iii) Attrition, which is a measure of the people successfully converted but not retained for converting again; (iv) Loyalty, which is a measure of the number of visits likely to be made by any visitor during the period as a visitor; (v) Frequency, which measures the activity generated by a visitor on a web site in terms of the time between visits, in the unit of days between visits; and (vi) Recency, which is the number of days passed after the last visit/purchase.

One of the important parameters of this science is the Decision Support System (DSS), which is a conceptual framework for the whole process of supporting the managerial decision making, based on modeling the problems and employing the quantitative models for analyzing and finding the solution. Business intelligence (BI), a subset of the DSS, is a very broad term which includes the architectures, tools, databases, applications, and methodologies. Another parameter, Business Analytics (BA) is a subset of the BI, which applies the models directly to the business data, and also helps in reaching the strategically important decisions. The term, Web Analytics (WA) is a subset of BA, and is based on the application of business analytics activities to Web based processes like e-commerce. The topic of BA has drawn the interest of various workers, and some books (Bartlett, 2013; Davenport and Jeanne, 2007; McDonald and Tina, 2007; Stubbs, 2011; Ranadive, 2006; and Zabin and Gresh, 2004), and comprehensive reviews (Davenport, 2006; Pfeffer and Robert, 2006; Davenport and Jeanne, 2005; Bonabeau, 2003; and Davenport et al, 2001) are available on various aspects of the subject.

### **Companies with a Large Number of Customers**

Companies which have direct contact with a large number of customers and a growing number channel-oriented applications like e-commerce support or call center support, have to face a new data management challenge of finding the effective way of the integration of all the enterprise applications in real time. In addition, based on learning from the past, forecast has to be made of the future, which is done by adopting Business Intelligence (BI) tools and systems. The importance of enforcing achievements of the goals defined by the business strategies through business intelligence concepts has now been well understood by the business houses, which describes the insights on the role and requirement of the real time BI by analyzing the business needs.

**Business Intelligence (BI)**

There are slightly varying definitions of the BI, some of them being like - (i) Business intelligence is the process, involving collection of the large amounts of data, analyzing that data, and preparing a high level set of reports, which condense the essence of that data into the basis of the business actions desired to be taken, and thus helps the management to make the fundamental daily business decisions; (ii) BI is the way and method of improving business performance by providing powerful assistance to the executive decision makers, by providing the actionable information. The BI tools are considered as the technology, which enables the efficiency of business operation by providing an increased value to the enterprise information, and also the manner in which this information should be utilized; (iii) BI is the process of collection, treatment and diffusion of information with an objective of reducing the uncertainty in the making of all strategic decisions.; and (iv) BI is a business management term used to describe applications and technologies used together, and provides access to analyze data and information about an enterprise, and thus helps them in making better informed business decisions. The topic of BI has drawn the attention of various workers in the last decade, and books (Stackowiak et al, 2007; Berson et al, 2002; and Malhotra, 2000), a comprehensive review (Adelman et al, 2002) and conference proceedings (Cui et al, 2007; Tvrdikova, 2007; Nguyen et al, 2005; and Siefert and Schiefer, 2005) are available on the various issues of interesting subject.

The main components of BI are the information delivery, reporting, statistical analysis, adhoc analysis, and predicative analysis. Business Intelligence is a technology and not just an integrated solution for companies, within which the business requirement is the important key factor driving the technology innovation. Recently, some other interesting investigations (Samuel and Guclu, 2008; Metzler, 2007; and Sanders, 2011) have been reported on the techniques related to the topic of business studies. Samuel and Guclu, (2008) have suggested a Real options approach for entrepreneurs making decisions under uncertainty. Metzler (2007) has discussed the wide area networks by optimizing the performance for the branch offices. Sanders (2011) has suggested and discussed in detail business process Reengineering and Knowledge value added approach.

**PREDICTIVE ANALYTICS**

Predictive analytics is the branch of data mining dealing with the prediction of future probabilities and trends, which are used to find the probable future outcome of an event or the likelihood of the occurrence of a situation. This technique is used to automatically analyze the large amounts of data including clustering, decision trees, market basket analysis, regression modeling, neural nets, genetic algorithms, text mining, and decision analytics, with different variables. The technique's core element is the predictor, a variable, which can be measured for an individual/entity for predicting the future behavior. For example, a life insurance company can consider age and health history, when issuing a policy to determine an applicant's risk factor.

Predictive Analytics is a science based on a set of business intelligence technologies, uncovering the relationships and patterns within large volumes of data, which can be used for predicting the behavior and the events. This technique is forward looking, and uses the past events to anticipate the future.

This is distinguished from the other BI technologies in the sense that whereas they are deductive in nature, and used for validating their hypotheses, this is inductive in nature, which is able to pull out the relationships and patterns.

The roots of the data mining lie in the predictive analytics. Still, the terms data mining and data extraction are often confused with each other in the corporate world. Data mining carries more meaning and importance than the data extraction, since it also involves the extraction of hidden predictive information from large databases. Data mining is also termed as the knowledge discovery in databases, by using data computational techniques from statistics and pattern recognition. Also, data extraction is the process of pulling data from one data source and loading the same into a targeted database; and pulling data from source or the legacy system and loading data into standard database or data warehouse. A predictive analytical model is based on the data mining tools and techniques. The Data mining tools extract the data by accessing massive databases and then processing the data with advanced algorithms to find hidden patterns and the predictive information.

Predictive analytics has to take both, the microscopic and telescopic views of the data, and use them to help the organizations to see and analyze the minute details of the business, and to plan for the future. Though the traditional BI tools are quite useful, they are based on certain created assumptions, and then finding if the statistical patterns do match these assumptions, and in this manner are quite limited in applications. However, the predictive analytics also discovers the previously unknown data; and then finally looks for patterns and associations for all places and all times.

### **Trends in Predictive Analytics**

It is a bit surprising that though a substantial increase in interest has been shown in the Predictive Analytics by the BI community, comparatively very few organizations have started using it. It is reported that ~ 40% organizations have implemented the predictive analytics as a mature tool by using the well defined processes, with good degrees of success, enabling them to evaluate and improve their modeling efforts in a continuous manner.

### **Predictive Analytics Software**

There are many Predictive Analytics Softwares available e.g. IBM/SPSS and SAS provide Integrated Analytic Workbenches, which provide support for Graphical modeling, automated testing, and text analytics. Also, the Industry standard methodology called the Cross Industry Standard Process is used for the Data Mining (CRISP-DM).

### **Process of Predictive Modeling**

The predictive modeling involves various steps: (i) Project Definition, which defines the business objectives and the desired results, (ii) Data Understanding, which analyses the source data to determine the appropriate data, and model building approach, (iii) Data Preparation, which does the selection, extraction and transformation of the data to create the required models, (iv) Model Building, which creates and validates models by evaluating them, (v) Deployment, which applies the model results to business decisions or processes, and (vi) Model Management, which manages models for improving performance, controlling access, and promoting the reuse.

**Predictive Modeling Methods**

In an ideal situation, a company has to look into the future and take appropriate actions before the customers get interested in the competitor companies. So, the company has to build a predictive model by using three predictor parameters: frequency of use, personal financial situations, and lower annual percentage rate (APR) offered by competitors, and then considering their combination.

The Analysts devise the specific models by using different techniques e.g. neural networks, decision trees, and linear regression. This is not just a routine technique, and requires skill in creating effective analytic model by choosing the models and algorithms suitable for use in the particular case, as the success depends on their proper choice. It is now the common practice to apply multiple models to a problem, and find the best working combination. The technique is so advanced, that even the persons without the specialization are able to design quite effective analytic models.

The final model to be employed is based on a linear combination of the candidate models. The effectiveness of the technique is increased because of the fact that after the finalization of the modeling data sets, the maximum incremental gain is achieved by fine tuning the training parameters, and combining the predictions from the multiple algorithms.

Prediction accuracy is evaluated based on the following objective function for the model to minimize

$$\sqrt{1/n \cdot \sum_i^n \{\log(pred_i + 1) + \log(act_i + 1)\}^2} \quad \text{---- (2),}$$

where *i* is a member, *n* is the total number of members, *pred<sub>i</sub>* is the predicted number of days spent in the organization for member *i* in the specific period, and *act<sub>i</sub>* is the number of days actually spent in the organization for member *i* in the specific period. Depending upon the size of the organization, the number *i* can be really large. Also, the choice of the specific period is very important from the point of view of the analysis.

To measure the matching of a predicted parameter with the corresponding measured parameter, the directional relative error (DRE) may be used, which is given by:

$$DRE = \frac{\{\text{predicted parameter} - \text{measured parameter}\}}{\min \{\text{measured parameter}, \text{predicted parameter}\}} \quad \text{---- (3).}$$

So, a value of zero implies a perfect prediction. Usually, a value of DRE between 0.1 – 0.2 is achievable by the business experts. However, after 2-3 iterations, a value close to zero is achieved.

**Commercially Available Predictive Analytics**

The importance and potential of the Predictive Analytics have grown so much that many commercial vendors are in the field prepared to do this work for the companies. Some of these are mentioned here: (i) SAS, which is one of the leaders in predictive analytics, engaged in making data mining and creating predictive analytic tools; (ii) SPSS Inc.,

which is also one of the leaders in providing predictive analytics software and solutions; (iii) Insightful Corporation, which is a supplier of software and services for statistical data analysis, data mining of numeric, and text data; (iv) StatSoft Inc., which is a global provider of analytic software, with its product named Statistica providing comprehensive array of data analysis, data management, data visualization, and data mining procedures; (v) Knowledge Extractions Engines (KXEN), which provide the KXEN Analytic Framework in the form of a suite of predictive and descriptive modeling engines for creating analytic models; (vi) Angoss Software Corporation, which makes products for providing information on customer behavior and marketing initiatives to help in the development of business strategies; (vii) Fair Isaac Corporation, which is the leading provider of credit scoring systems; and (viii) IBM, which offers predictive analytics tools like - DB2 Intelligent Miner for Data, that is a predictive analytical tool useful to gain new business insights, and to harvest valuable business intelligence from the enterprise data.

## **CONCLUSION**

It is now well understood that transaction oriented information systems have to be employed in every major industry or the corporation for the optimum functioning. However, the analytically oriented systems are required by them in order to remain competitive, which can revolutionize the company's ability to utilize the already owned information. The (BI) has evolved very strongly to rely increasingly on real time data. The BI systems start the actions for the systems automatically, based on rules with reference to supporting the several business processes, and in turn, these analytical systems derive insight from the wealth of available data, for delivering the conclusive information. It is now necessary to do the business analysis, and perform the actions in response to the analysis of results, and also instantaneously change the parameters of business processes. As discussed here, the BA is the practice of iterative, methodical exploration of an organization's data based on the statistical analysis, which is used by the companies committed to data driven decision-making. Because of the increasing popularity of business analytics, BI application vendors are including some BA functionality in their products e.g. Data Warehouse Appliance vendors have started to embed BA functionality within the appliance. Some other major enterprise system vendors are also putting more analytics into the memory, and it is expected to shorten the time between a business event and the decision-making. Thus, it can be safely concluded that the field of Modeling and Technical Analysis of Electronics Commerce and Predictive Analytics is on a solid footing, and also evolving fast.

## **ACKNOWLEDGEMENTS**

The author is grateful to the Dr. Nand Kishore Garg, Chairman, Maharaja Agrasen Institute of Technology, GGSIP University, Delhi for providing the facilities for carrying out this research work, and also for his moral support. The author is thankful to Dr. M. L. Goyal, Director, for encouragement. Thanks are also due to Dr. V. K. Jain, Deputy Director for his support during the course of the work. The author is thankful to Prof. V. K. Tripathi, Department of Physics, Indian Institute of Technology, Delhi for useful discussions and various suggestions which have significantly improved the presentation of the paper.



## REFERENCES

- Adelman S., Moss L., and Barbusinski L., I found several definitions of BI,DM Review (2002). Retrieved 17 August 2002 from [http://www.dmreview.com/article\\_sub.cfm?articleId=5700](http://www.dmreview.com/article_sub.cfm?articleId=5700).
- Bartlett R., (2013). A Practitioner's Guide To Business Analytics: Using Data Analysis Tools to Improve Your Organization's Decision Making and Strategy. McGraw-Hill. ISBN 978 0071807593.
- Berson A., Smith S., and Thearling K., (2002) ,Building Data Mining Applications for CRM, Tata McGraw Hill, 540 pages.
- Bonabeau E., (2003). Don't Trust Your Gut. Harvard Business Review, 81(5), 116-123.
- Cui Z., Damiani E., and Leida M., (2007). Benefits of Ontologies in Real Time Data Access', Digital Ecosystems and Technologies Conference (DEST), 07, 392-397.
- Davenport Thomas H. (2006). Competing on Analytics. Harvard Business Review, <http://hbr.org/2006/01/competing-on-analytics/ar/1>.
- Davenport Thomas H. and Jeanne G. Harris, (2007). Competing on Analytics: The New Science of Winning, Harvard Business School Press.
- Davenport Thomas H. and Jeanne G. Harris, (2005), Automated Decision Making Comes of Age. MIT Sloan Management Review, <http://sloanreview.mit.edu/article/automated-decision-making-comes-of-age/>.
- Davenport Thomas H., Jeanne G. Harris, David W. De Long, and Alvin L. Jacobson. (2001), Data to Knowledge to Results: Building an Analytic Capability, California Management Review, 43, No.2, 117-138.
- Malhotra Y., (2000). Information Management to Knowledge Management: Beyond Hi-Tech Hidebound Systems, in Srikantaiah, T. K. and Koenig, M.E.D. (Eds.) Knowledge Management, Medford, NJ., Information Today, 37-61.
- McDonald M. and Tina N., (2007), Creating Enterprise Leverage: The 2007 CIO Agenda. Stamford, CT: Gartner, Inc.
- Metzler J., (2007), Wide Area Networks: Optimizing Performance for your Branch Offices. Business Communication Review, 37, 52-55. Retrieved January 12, 2011,from <http://libproxy.nps.edu/login?url=http://proquest.umi.com.libproxy.nps.edu/pqdweb?did=1309671431&sid=1&Fmt=4&clientId=11969&RQT=309&VName=PQD>.
- Nguyen Tho M., Schiefer J., and Min Tjoa, A., (2005), Data warehouse design 2: Sense and response service architecture (SARESA): an approach towards a real-time business intelligence solution and its use for a fraud detection application, Proceedings of the 8th ACM international workshop on Data warehousing and OLAP, ACM Press. DOLAP '05 (2005).
- Pfeffer J. and Robert I. Sutton, (2006), Evidence-Based Management. Harvard Business Review, 62-74.
- Ranadive Vivek, (2006) , The Power to Predict: How Real Time Businesses Anticipate Customer Needs, Create Opportunities, and Beat the Competition. McGraw-Hill. (2006-01-26).
- Samuel, S. R., and Guclu, A., (2008), A Real options Approach for Entrepreneurs making Decisions under uncertainty. Issues in Innovation, Retrieved from [http://issuesininnovation.org/Documents/2\(1\),42](http://issuesininnovation.org/Documents/2(1),42).
- Sanders Aaron D., (2011), WAN Optimization: A Business Process Reengineering and Knowledge Value added Approach M. Sc. (Information Systems Technology)

- Thesis, Naval Postgraduate School Monterey, California, USA, 73 pages.
- Stackowiak, R., Rayman, J. and Greenwald, R., (2007), Oracle Data Warehousing and Business Intelligence Solutions, Wiley Publishing, Inc, Indianapolis, 386 pages.
- Stubbs E., (2011), The Value of Business Analytics. John Wiley and Sons, ISBN: 978-1-118-01239-0, 322 pages.
- Suefert A. and Schiefer J., (2005), Enhanced Business Intelligence- Supporting Business Processes with Real-Time Business Analytics, Proceedings of the 16th international workshop on Database and Expert System applications-DEXA'05 (2005). Retrieved 19 June 2006 from [www.ieee.org](http://www.ieee.org).
- Tvrdikova M., (2007), Support of Decision Making by Business Intelligence Tools', Computer Information Systems and Industrial Management Applications, 6th International Conference, CISIM 07, 368.
- Zabin J. and Gresh B., (2004), Precision Marketing. The New Rules for Attracting, Retaining, and Leveraging Profitable Customers, John Wiley and Sons, 06-Feb-2004, 256 pages.