

2013

Research and Pedagogy in Business Analytics: Opportunities and Illustrative Examples

Ramesh Sharda

Daniel Asamoah

Wright State University - Main Campus, daniel.asamoah@wright.edu

Natraj Ponna

Follow this and additional works at: https://corescholar.libraries.wright.edu/infosys_scm



Part of the [Management Information Systems Commons](#), [Operations and Supply Chain Management Commons](#), and the [Social Media Commons](#)

Repository Citation

Sharda, R., Asamoah, D., & Ponna, N. (2013). Research and Pedagogy in Business Analytics: Opportunities and Illustrative Examples. *Journal of Computing and Information Technology*, 21 (3), 171-183.
https://corescholar.libraries.wright.edu/infosys_scm/4

This Article is brought to you for free and open access by the Information Systems and Supply Chain Management at CORE Scholar. It has been accepted for inclusion in ISSCM Faculty Publications by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu, library-corescholar@wright.edu.

Research and Pedagogy in Business Analytics: Opportunities and Illustrative Examples

Ramesh Sharda, Daniel Adomako Asamoah and Natraj Ponna

Institute for Research in Information Systems, Spears School of Business, Oklahoma State University, USA

In recent times, business analytics and big data have gained momentum both in industry practice and academic research. The objective of this paper is to provide both a research and teaching introduction to business analytics in the context of both current and prospective perspective of the business analytics domain. It begins by providing a quick overview of the three types of analytics. To assist the future analytics professionals, we identify, group and discuss nine different participants of the analytics industry into clusters. We then include a brief description of some current research projects under way in our team. We also note some research opportunities in Big Data analytics. The paper also concludes with a discussion of teaching opportunities in analytics.

Keywords: business analytics, Big Data, social media research, athletic injuries, healthcare

1. Introduction

Although many authors and consultants have defined it slightly differently, analytics can be viewed as the process of developing recommendations for actions or actionable decisions based upon insights generated from historical data. The Institute for Operations Research and Management Science (INFORMS) has created a major initiative to organize and promote analytics. According to INFORMS, analytics represents the combination of computer technology, management science techniques, and statistics to solve real problems. Of course, many other organizations have proposed their own interpretations and motivation for analytics. For example, SAS Institute Inc. proposed eight levels of analytics that begin with standardized reports from a computer system. These reports provide

a sense of what is happening with an organization. Additional technologies have enabled us to create more customized reports that can be generated on an ad hoc basis. The next extension of reporting takes us to online analytical processing (OLAP) type queries that allow users to dig deeper and determine specific sources of concern or opportunities. Technologies available today can also automatically issue alerts for decision makers when the performance issues warrant such alerts. At a consumer level, we see such alerts for weather or other issues. But similar alerts can also be generated in specific settings, for example, when the sales fall above or below a certain level within a certain time period, or when the inventory for a specific product is running low. All of these applications are made possible through analysis and queries on data being collected by an organization. The next level of analysis entails statistical analysis to better understand patterns, which can be used to develop forecasts or models for predicting how customers might respond to a specific marketing campaign or ongoing service/product offering. When an organization has a good view of what is happening and what is likely to happen, it can also employ other techniques to make the best decisions under the circumstances. These eight levels of analytics are described in more detail in a white paper by SAS. (http://www.sas.com/news/sascom/analytics_levels.pdf) [13].

Instead of eight levels, INFORMS defines three categories of analytics, based on the idea of looking at all the data to understand what is happening, what will happen, and how to make

the best of it. These three levels are identified (<http://www.informs.org/Community/Analytics>) as Descriptive, Predictive, and Prescriptive [5]. Figure 1 presents a graphical view of these three levels of analytics. The interconnected circles show that there is some overlap across these three types of analytics. A more detailed description of each of these categories follows.

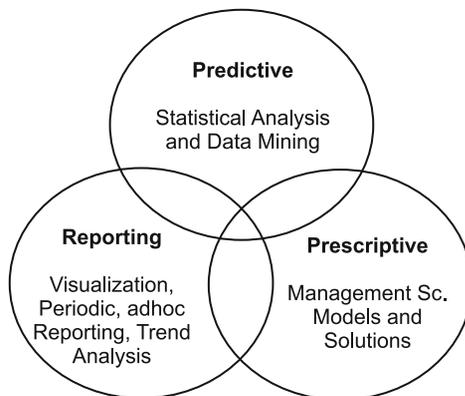


Figure 1. Three types of analytics.

Descriptive or reporting analytics refers to knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences. This involves consolidation of data sources and availability of all relevant data in a form that enables appropriate reporting and analysis. From this data infrastructure, we can develop appropriate reports, queries, alerts, and trends using various reporting tools and techniques. In our view, this has been the key focus of business intelligence (BI). A significant technology that has become a key player in this area is visualization. Using the latest visualization tools in the market, we can now develop powerful insights into the operations of an organization.

Predictive analytics aims to determine what is likely to happen in the future. This analysis is based on statistical techniques as well as other more recently developed techniques that fall under the general category of data mining. The goal of these techniques is to be able to predict if the customer is likely to switch to a competitor (“churn”), what a customer is likely to buy next and in what quantity, what promotion a customer would respond to, whether this

customer is a credit risk, etc. A number of techniques are used in developing predictive analytical applications including various classification algorithms. For example, we can use classification techniques such as decision tree models and neural networks to predict how well a motion picture would do at the box office. We can also use clustering algorithms for segmenting customers into different clusters to be able to target specific promotions to them. Finally, we can use association mining techniques to estimate relationships between different purchasing behaviors. That is, if a customer buys one product, we can predict other items the customer is likely to purchase. Such analysis can assist a retailer in recommending or promoting related products. For example, any product search on Amazon.com results in the retailer also suggesting other similar products that a customer may be interested in.

The third category of analytics is termed Prescriptive analytics. The goal of prescriptive analytics is to examine current trends and likely forecasts and use that information to make decisions. This group of techniques has historically been studied under the umbrella of operations research or management sciences and is generally aimed at optimizing the performance of a system. The goal here is to provide a decision or a recommendation for a specific action. These recommendations can be in the forms of a specific yes/no decision for a problem, a specific amount (say, price for a specific item or airfare to charge), or even a complete set of production plans. The decisions may be presented to a decision maker in a report, or may directly be used in an automated decision rules system (e.g., in airline pricing systems). Thus, these types of analytics can also be termed Decision or Normative Analytics.

Over the years, the use of the analytics and its related fields has grown tremendously and now there are wide varieties of industry participants that use analytics. They range from industries that provide data infrastructure, data warehousing solutions, middleware, data aggregation, and analytics software developers to analytics user organizations and academics. Research across all three categories of analytics also continues to flourish. Technology enhancements have enabled application of analytic techniques to gain insights from large amounts of both

structured and unstructured data. This forms the basis of Big Data analytics.

This paper is organized as follows. In Section 2, we identify various participants in the analytics industry and group them into clusters. Section 3 discusses research in predictive analytics and introduces an example of a predictive analytics project the authors have worked on recently. Section 4 presents a discussion on research options in Big Data analytics. This section also includes a brief description of one of our research projects that employs Big Data analytics. Lastly, we discuss the role of academic programs and modules in advancing the cause of both research and practitioner development of business analytics. We note that this paper is an extension of a paper presented at the 35th International Conference on Information Technology Interfaces, Dubrovnik, Croatia, June 2013 [18].

2. Analytics Industry Participant Clusters

This section is aimed at identifying various analytics industry players by grouping them into various sectors. We note that the list of company names included is not exhaustive. These merely reflect our own awareness and mapping of companies' offerings in the analytics industry space. Additionally, the mention of a company's name or its capability in one specific group does not mean that the specific activity is the only offering of that organization. We use these names simply to illustrate our descriptions of sectors.

Many other organizations exist in this industry. Our goal is not to create a directory of players or their capabilities in each space, but to illustrate to students considering a career in analytics that many different options exist for playing in the analytics industry. One can start in one sector, and move to another role altogether. We will also see that many companies play in multiple sectors within the analytics industry and thus offer opportunities for career transitions within the field both horizontally and vertically.

Figure 2 presents the analytic industry clusters categorized into nine key sectors or clusters, forming Analytics Ecosystem. The first five clusters can be broadly termed technology providers or accelerators. Their primary revenue comes from developing technology, solutions, and training which enables the user organizations to employ these technologies in the most effective and efficient manner. The accelerators include academics and industry organizations whose goal is to assist both technology providers and users.

2.1. Data Infrastructure Providers

This cluster includes all major organizations that provide the hardware and software infrastructures targeting the basic foundation for the data management solutions. Some of the most obvious players in this cluster: companies providing hardware infrastructure for database computing like IBM, Dell, HP and Oracle; storage solution providers like EMC and NetApp;

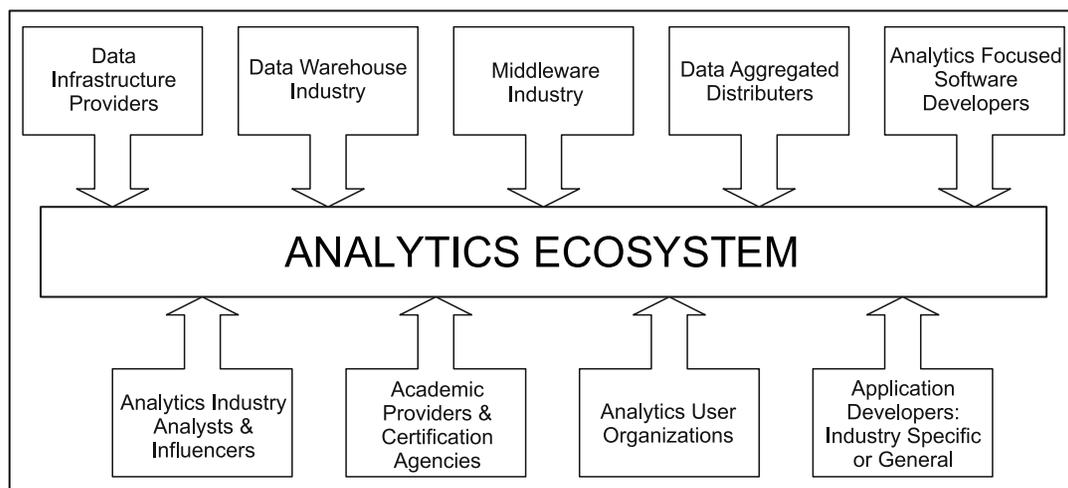


Figure 2. Analytics industry clusters.

companies providing indigenous hardware and software platforms such as IBM and Teradata; data solution providers offering hardware and platform independent database management systems like the SQL Server family of Microsoft; specialized integrated software providers such as SAP; companies like Amazon and Salesforce.com providing full data storage solutions through cloud computing; the recent crop of companies in big data space like Cloudera, Hortonworks, Alteryx, Pivotal and Hadapt; and companies providing infrastructure services and training to create Big Data platform such as Hadoop Clusters, MapReduce, NOSQL.

2.2. Data Warehouse Industry

Companies with a data warehousing focus provide technology and services aimed towards integrating data from multiple sources thus enabling organizations to derive and deliver value from its data assets. They form the backbone for all the players in the analytics industry. Many companies in this space include their own hardware to provide efficient data storage, retrieval, and processing. Recent developments in this space include performing analytics on the data directly in memory. Companies such as IBM, Oracle, and Teradata are major players in this arena.

2.3. BI Platforms/Middleware Industry

The general goal of middleware industry is to provide easy to use tools enabling reporting or descriptive analytics on the enterprise-wide data. This forms a core part of BI or analytics employed at organizations. Examples of companies in this space include Microstrategy, Plum, Oracle–Hyperion, SAP–Business Objects, IBM-Cognos and many others.

2.4. Data Aggregators/Distributors

These companies typically focus on a specific industry sector for specialized data collection, aggregation and distribution through their niche platforms and services for data collection. They build their relationships by providing the data to other industry players and enabling them to employ analytics. Some of the example companies

include: Nielsen, which provides data sources to their clients on customer retail purchase behavior; Experian, which provides data on each household in the US; Omniture, which has developed technology to collect web clicks and share such data with their clients; and Google, which compiles data for individual websites and makes a summary available through Google Analytics services.

2.5. Analytics Focused Software Developers

Companies in this category have developed analytics software for general use with data that has been collected in a data warehouse or is available through one of the platforms identified earlier (including Big Data). It can also include inventors and researchers in universities and other organizations that have developed algorithms for specific types of analytics applications. We can identify major industry players in this space along with the three types of analytics outlined earlier.

2.5.1. Reporting/Analytics

Reporting or descriptive analytics is enabled by the tools available from the Middleware industry players or unique capabilities offered by focused providers. For example, Microsoft's SQL Server BI tool kit includes reporting as well as predictive analytics capabilities. On the other hand, specialized software is available from companies such as Tableau for visualization. SAS also offers a Visual Analytics tool for similar capacity. Likewise, there are many open source visualization tools as well.

2.5.2. Predictive Analytics

Perhaps the biggest recent growth in analytics has been in this category, resulting in a large number of companies that focus on providing both technical and management solutions for predictive analytics. Many statistical software companies such as SAS, SPSS, etc. embraced predictive analytics early on and developed the software capabilities as well as industry practices to employ data mining techniques as well

as classical statistical techniques for analytics. IBM-SPSS Modeler from IBM and Enterprise Miner from SAS are some of the examples of tools used for predictive analytics. Other players in this space include KXEN, Statsoft, Salford Systems, and scores of other companies that may sell their software broadly or use it for their own consulting practices. Two open source platforms (R and RapidMiner) have also emerged as popular industrial strength software tools for predictive analytics and have companies that support training and implementation of these open sources tools.

2.5.3. Prescriptive Analytics

Software providers in this category offer modeling tools and algorithms for optimization of operations tagged as management science/operations research (MS/OR) software. This field has had its own set of major software providers. IBM, for example, has classic linear and mixed integer programming software, IBM-ILOG, which provides prescriptive analysis services. Analytics providers such as SAS have their own OR/MS tools – SAS/OR. Other major players in this domain include companies such as AIMMS, AMPL, Frontline, GAMS, Gurobi, Lindo Systems, Maximal, and many others.

Of course, there are many techniques that fall under the category of prescriptive analytics, each having their own set of providers. For example, simulation software is provided by major companies like Rockwell (ARENA) and Simio. Similarly, Frontline offers tools for optimization with Excel spreadsheets. Decision analysis in multi-objective settings can be performed using tools such as Expert Choice. There are also tools from companies such as Exsys, XpertRule, and others for generating rules directly from data or expert inputs. Some new companies are evolving to combine multiple analytics models in the Big Data space. For example, Teradata Aster includes its own predictive and prescriptive analytics capabilities in processing big data streams.

2.6. Application Developers: Industry Specific or General

The organizations in this group use their analytical expertise, focusing on using solutions avail-

able from the data infrastructure, data warehouse, middleware, data aggregators, and analytics software providers to develop custom solutions for a specific industry. Thus, this industry group makes it possible for the analytics technology to be truly useful. Most major analytics technology providers like IBM, SAS, Teradata, etc., clearly recognize the opportunity to connect to a specific industry or client and offer analytic consulting services. Companies that have traditionally provided application/data solutions to specific sectors are now developing industry specific analytics offerings. For example, Cerner provides electronic medical records (EMR) solutions to medical providers and their offerings now include many analytics reports and visualizations. Similarly, IBM offers a fraud detection engine for the health insurance industry, and is working with an insurance company to employ their famous Watson analytics platform in assisting medical providers and insurance companies with diagnosis and disease management. Another example of a vertical application provider is Sabre Technologies, which provides analytical solutions to the travel industry including fare pricing for revenue optimization, dispatch planning, etc.

This cluster also includes companies that have developed their own domain specific analytics solutions and market them broadly to a client base. For example, Axiom has developed clusters for virtually all households in the US based upon all the data they collect about households from many different sources. Credit score and classification reporting companies (Equifax, Experian, TransUnion, etc.) also belong to this group. Demandtec (now owned by IBM) provides pricing optimization solutions in retail industry. This field represents an entrepreneurial opportunity to develop industry specific applications. Examples of such companies and their activities are: Sense Networks, which employs location data for developing user/group profiles, X+1 and Rapleaf, which profile users on the basis of email usage, Bluecava, which aims to identify users through all device usage, and Simulmedia, which targets advertisements on TV on the basis of analysis of a user's TV watching habits.

2.7. Analytics User Organizations

Clearly, this is the economic engine of the whole analytics industry. If there were no users, there would be no analytics industry. Organizations in every other industry, size, shape, and location are using analytics or exploring use of analytics in their operations, trying to gain/retain a competitive advantage. These include organizations in private sector, government, education, military, etc. Examples of uses of analytics in different industries are too numerous to list here. Also, specific companies are not identified in this section. Rather, the goal is to see what type of roles analytics professionals can play within a user organization.

Of course, the top leadership of an organization, (Chief Information Officer, etc) is critically important in applying analytics to its operations. Although not enough senior managers seem to subscribe to this view, the awareness of applying analytics within an organization is growing everywhere. Major organizations in every industry that we are aware of are hiring analytical professionals under a variety of titles; these analytics professionals are eventually moving into management positions and driving business decision making using insights derived by employing analytics.

2.8. Analytics Industry Analysts and Influencers

This cluster of analytics industry includes three types of organizations or professionals. The first group is the set of professional organizations that provide advice to the analytics industry providers and users. Their services include marketing analyses, coverage of new developments, evaluation of specific technologies, and development of training and white papers. Examples of such players include organizations such as the Gartner Group, The Data Warehousing Institute, Forrester, McKinsey, and many of the general and technical publications and websites that cover the analytics industry.

The second group includes professional societies or organizations that also provide some of the same services, but are membership-based. For example, INFORMS, a professional organization, has now focused on promoting analytics. Special Interest Group on Decision

Support Systems (SIGDSS), a subgroup of the Association for Information Systems, also focuses on analytics. Most of the major vendors (e.g., Teradata, SAS, etc.) also have their own membership-based user groups promoting the use of analytics.

A third group of analytics industry analysts is what we call analytics ambassadors, influencers, or evangelists. They have presented their enthusiasm for analytics through their seminars, books, and other publications. Illustrative examples include Steve Baker, Tom Davenport, Charles Duhigg, Wayne Eckerson, Bill Franks, Malcolm Gladwell, Claudia Imhoff, Bill Inman, and many others.

2.9. Academic Providers and Certification Agencies

Academic providers and certificate agencies play a major role in fueling the analytics industry with the necessary human resources by preparing students with current analytics expertise. This cluster, made up of different programs in business schools (information systems, management, marketing etc.), hence represents the academic programs that prepare professionals for the analytics industry. Other academic programs such as computer science, statistics, mathematics and industrial engineering offer courses that help train students for the analytics industry.

The certification programs established by technology providers such as IBM, Microstrategy and SAS have a paramount aim of giving students the needed set of tools that are relevant to solving real world problems in the industry. Accompanying exercises and laboratory assignments are mostly adapted from actual implementations in businesses. For instance, INFORMS has just introduced a Certified Analytics Professional (CAP) certificate program that is aimed at testing an individual's general analytics competency.

Collaborative initiatives also exist between formal academic institutions and technology providers that help students develop strong theoretical background of analytic concepts, sharpen their problem formulation techniques and help them solve real world problems. Among other targets, the collaborative initiatives between academic institutions and technology providers are

aimed at helping students develop better tools, procedures and algorithms through research activities. The outputs of such research activities are then translated into providing both rigorous and practically sound solutions to real world problems encountered by businesses.

3. Predictive Analytics Research Examples

Research in applying predictive analytics has been growing exponentially over the last few years. The first wave of the research focused on developing algorithms in major categories of predictive analytics: classification, clustering, association mining, etc. These included decision trees, apriori algorithms, neural networks, and hundreds of other algorithms and variants. These algorithms have been applied in thousands of exploratory projects. A catalog of such research projects will fill a book, hence we will only illustrate these opportunities by providing a brief description of one recent project in applying predictive analytics in healthcare and sports.

3.1. Analyzing Athletic Injuries

Any athletic activity is prone to injuries and if the injuries are not handled properly, adverse performance-related impact to athletic teams

could result. Using analytics to understand injuries can help in deriving valuable insights that would enable the coaches and the medical team to manage the team composition, understand the player profiles and ultimately aid in better decision making on the availability of players.

In an exploratory study, injuries related to American Football were analyzed using reporting and predictive analytics by our research team. The project followed the CRISP-DM methodology to understand the business problem of making recommendations on managing injuries, understand the various data elements collected about injuries, cleaning the data, performing relevant imputations for the missing data, developing visualizations to draw various inferences, building predictive models to analyze the injury healing time period and drawing sequence rules to predict the relationship among the injuries and the various body parts that were afflicted with injuries.

The injury data set consisted of over 560 football injury records that include injury-specific variables: body part, body site, laterality, action taken, severity, injury type, injury start date and healing date. It also included player identity and other sport specific variables such as position played, activity, onset, and game location. Based on the injury start date and the healing date, healing time was calculated for each record and classified into different sets of time

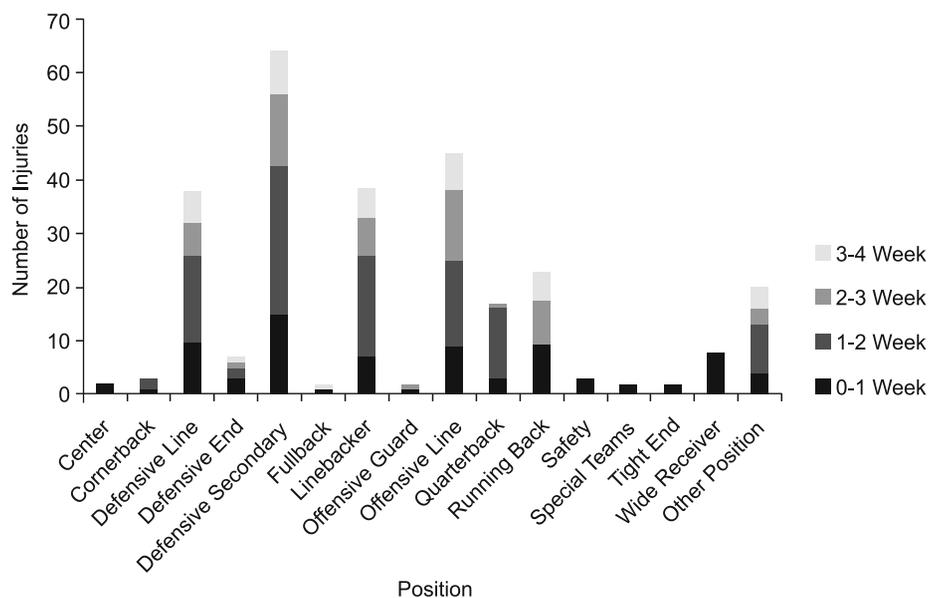


Figure 3. Injury recovery period at various positions.

periods: 0-1 month, 1-2 months, 2-4 months, 4-6 months and 6-24 months. The overall injury records with valid recovery time limited the final data available for modeling to 384 records.

We first began with descriptive analytics. Various visualizations were built to draw inferences, depicting the healing time period associated with player’s positions, severity of injuries and the healing time period, treatment offered and the associated healing time period, major injuries afflicting body parts, type of field and game etc. Some of these are shown in the figures below.

Figure 3 depicts the period of recovery from injuries sustained by players at various positions. The injury recovery variable ranges are 0 to 1 week, 1 to 2 weeks, 2 to 3 weeks and 3 to 4 weeks. It is realized that fifty percent of injuries are resolved in less than 2 weeks. Also, defensive positions are more likely to sustain injuries as compared to other positions.

The visualization in Figure 4 indicates the various motions that lead to players’ injuries. The players are further segmented into different playing positions. It is seen from Figure 4 that tackling is the main cause of injuries in defensive players, such as the defensive secondary and linebacker positions while running is the main

cause of injuries in offensive players, such as the running back and wide receiver positions.

Our analysis included many more visualizations that enabled better and richer understanding of the problem domain and the data. We then began to apply predictive modeling methods to be able to understand the injury recovery rates. IBM SPSS Modeler was used for predicting each of the healing time period categories. The data was divided into training and validation data sets and several predictive models using techniques like CHAID, CRT and Neural Networks were built. Of all the techniques, a neural network using multi-layer perceptron and a single hidden layer with 15 neurons yielded the best results.

Some of the predictor variables in the neural network model were: current status of injury, severity, body part, body site, type of injury, activity, event location, action taken and position played. The prediction results shown in Table 1 depict the accuracy of the model in predicting how long an athlete remains injured once an injury is sustained. The relevant values are the diagonal figures which indicate the percent accuracy with which the recovery period for different periods of injuries could be predicted.

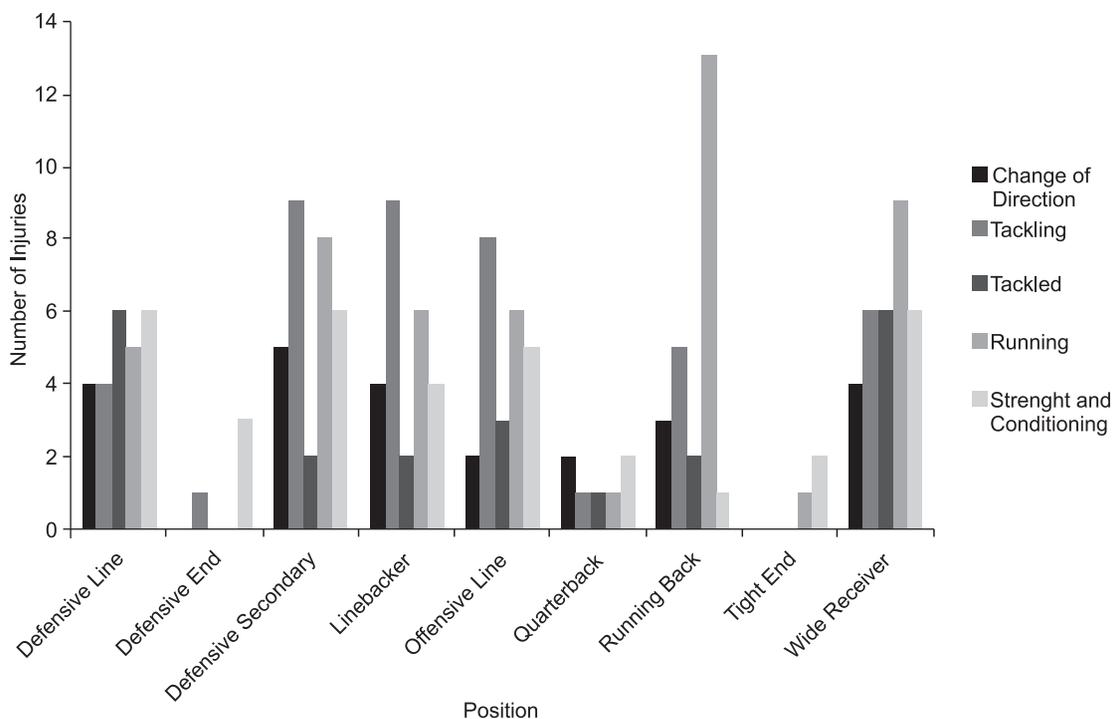


Figure 4. Motions at time of injury to various player positions.

For instance, with an accuracy of 75%, injuries that take 2 to 4 months to heal can be correctly predicted to take that long to heal based on the variables listed above in the prior sections. The overall percentage accuracy of the correct predictions of healing time period categories was 89.8%. The classification for the different healing periods was generated after initial analysis of the data set and in consultation with industry experts. The relevant injury period of interest with regards to college sports was chosen as a maximum of 2 years since most players only start active game time participation in the last two years of being on the football team. In this preliminary study, a classification period in months was chosen in order to give a broader level view of predicting injury recovery periods. In order to derive an in-depth understanding of the injury recovery pattern, further studies would drill down and consider injury recovery periods in weeks and days.

Observed (month)	Predicted (months)				
	0-1	1-2	2-4	4-6	6-24
0-1	97.3%	2.2%	0.5%	0.0%	0.0%
1-2	44.8%	55.2%	0.0%	0.0%	0.0%
2-4	18.8%	6.2%	75.0%	0.0%	3.9%
4-6	16.7%	0.0%	0.0%	83.3%	0.0%
6-24	0.0%	0.0%	25.0%	0.0%	75.0%

Table 1. Classifications for healing category.

Furthermore, we also performed a sequence analysis with injuries and body parts as precedents and antecedents respectively. The players are most likely to suffer the same injury again (repeated injuries) associated with same body part. It was also interesting to note that lower extremity injuries resulted in upper extremity injuries. Some of these include possible injuries afflicting the hand when a player had already had an injury to the ankle and injuries afflicting shoulder when a player has had an injury to the femur. For instance, with a support of 25% and confidence of 22.5%, a strain to the femur would result in a consequent shoulder injury.

Based on the analysis, many business recommendations were suggested which include the following: employing the services of healthcare specialist at the onset of the injury instead of

letting the training room staff screen the injured players; training players at defensive positions to increase their reflexive abilities in order to avoid getting injuries, and taking precautionary measures for the future injuries that a player might suffer based on results of the sequence analysis.

This research project illustrates an interesting application of predictive modeling methods in athletic health issues. The next section describes an emerging area of research in analytics.

4. Big Data Research Opportunities

The bottleneck to knowledge generation and scientific progress is now an issue of effective data analysis rather than data availability. Contemporary lifestyles of individuals and technological advancements in most parts of the world have resulted in the generation of huge amounts of data in astounding proportions every hour, in the form of emails, weblogs, tweets, sensor generated data, phone logs, GPS data, traffic data, video, and text. The data sources are a mixture of both structured and unstructured data formats. Hitherto, large volumes of structured data have been mainly collected, stored and mined using traditional and relational data warehousing technologies. However, the fast growth of data has resulted in a new breed of analytics called Big Data analytics. Big Data analytics supports large scale data analysis that focuses on large volumes of historical and near real-time data of different varieties and that needs to be analyzed at a fast pace.

At the core of Big Data research are business concepts, technologies and statistical techniques necessary to design highly scalable systems that can collect, process, store and analyze large volumes of both structured and unstructured data. One of the most popular platforms used to support processing of Big Data is an open source platform called Apache Hadoop. Apache Hadoop has two main basic components; Hadoop Distributed File System (HDFS) and MapReduce. HDFS, a distributed and scalable file system contributed by Google, supports the storage of data and computations on multiple commodity computers. MapReduce is a high level programming model that allows for parallel and distributed algorithms to be written for processing massive sets of data on a

cluster of computers. Besides these two components, the open source community has developed other sub-projects that help efficiently manage massive amounts of data. Some of these sub-projects are Hive, Pig, Mahout and Hbase. For instance, Hive is an SQL-like data warehouse platform that supports data operations such as data summarization and complex queries of data sets. The sub-projects may have different or over-lapping roles, yet they all are designed to facilitate efficient storage, manipulation and analysis of massive amounts of data. The Apache Hadoop platform is not a replacement for traditional relational database systems, but rather a complement that can better handle both structured and unstructured data sources.

Just like other streams in the data analytics field, research in Big Data analytics is applicable to almost all fields of practice and study. In areas like healthcare, logistics, marketing, retail and defense, it has become imperative to explore and analyze large amount of data in ways that supports effective decision making and strategic management. Such research and practice is growing rapidly. The highly scalable platform used for Big Data analytics is necessitated by three fundamental characteristics of data sources; volume, velocity and variety. Big Data research looks at how to analyze data in different domains with such characteristics and in a way that generates deeper knowledge and adds value to the decision making process in businesses. In the following, we describe one project on which we are currently working.

4.1. Analyzing the Role of Social Media in Healthcare Delivery

One of the nascent research streams in Big Data analytics is the role of electronic social media platforms in healthcare delivery. Early studies have established that factors such as evolving patient demographics, financial constraints and longer incidence of chronic diseases have enhanced the impact of social relationships in healthcare delivery [15]. Social relationships, which are associative bonds established between individuals with common aspirations and for the purpose of giving and receiving support, foster disease self-management programs that are vital to the management of chronic diseases.

Even though healthcare delivery and support has been traditionally made through channels

such as physicians, nurses and other clinical experts, the advent of the Web 2.0 platform has ushered an era where patients are beginning to inculcate new healthcare delivery strategies in the self-management of chronic diseases. In fact, some studies have suggested that traditional sources of health management information may no longer be enough in providing quality healthcare to patients [4]. Pharmaceutical companies and other healthcare-related businesses already market drugs and solicit consumer feedback from social networks about new drugs and disease management options.

Among a myriad of health related topics discussed on social media platforms is the management of chronic diseases such as diabetes, chronic obstructive pulmonary disease (COPD), major depressive disorder (MDD), epilepsy and cardiac failure [2]. These social media platforms offer social support and information to patients and caregivers. Social support has been identified as a major factor for improving health outcomes in patients who suffer from various types of illness [3][6]. The concept of social support, also sometimes referred to as peer support has been shown to not only help disease prevention, but also foster and promote general well-being and health [15][14][7]. The concept of social relationships and how they can be established on electronic social media platforms in the context of the provision of healthcare support is still at an early stage in academic research. More research needs to be conducted to establish the role and effectiveness of social media platforms in healthcare delivery.

Previous studies have mostly investigated efforts made by healthcare institutions in using social media as a tool for healthcare management and information delivery [9]. Armstrong and Powell (2009) assert that patients value the additional support they receive from online communities and perceive online sources of information provided by lay people as a complement to information provided by professional healthcare providers [1]. In addition, studies have shown that peer-to-peer support creates an enabling environment for disease self-management. However, very few studies have investigated how disease management and healthcare support are initiated by patients in a social media domain. Patients like to be in contact with others with similar problems in order to draw both emotional support and obtain information on how to manage diseases [1].

It has been established that social media platforms are a viable avenue for disease self-management. The question, however, still remains whether self-management programs through social media platforms are an effective means of managing chronic diseases. Disease self-management, especially for chronic diseases, can be an important aspect of the healthcare delivery process. Self-management programs are designed as a complement to other treatment methods to put the patients in the driving seat and empower them to take control of their health care status. The aim of a self-help program includes

1. teaching the appropriate use of medications,
2. evaluating methods and use of new treatment methods,
3. engaging in physical exercises that help decrease instance of fatigue, isolation and pain,
4. better ways of communicating with family, peers and clinical healthcare givers.

There is a need to investigate the effect of peer support and information exchange on social media in self-managing chronic disease conditions in an environment that is not directly mediated by a healthcare institution or any clinical stakeholder. Sarasohn-Kahn (2008) reported that, with respect to quality, expert information provided by a large collection of users on a social media platform is known to be at par with what would be provided by any single expert on a subject matter [12]. The same is true for clinical information provided by other users on a social networking platform. Hence, besides the healthcare advice from expert clinicians such as doctors, the advice and support provided on social media platforms by non-expert patients is useful to the achieving of the necessary health goals that need to be attained by patients.

As the focus of traditional health informatics shifts from health professionals to consumers, Big Data research assumes a vital role because it gives one the opportunity and flexibility to analyze extremely huge amounts of data rather than just a small sample and make inferences and decisions on the use and delivery of healthcare on social media platforms. Big Data research also delves into in-depth analysis and drawing of deeper conclusions on data generated from other domains of study.

5. Teaching Perspectives

In any knowledge intensive industry such as analytics, the fundamental strength comes from having students who are interested in the technology and choose that industry as their profession. Universities play a key role in making this possible. Academic programs that prepare professionals for the industry are increasing in number and type. These programs span various components of business schools such as information systems, marketing, management sciences, etc. Such programs are also emerging far beyond business schools to include computer science, statistics, mathematics, and industrial engineering departments across the world. These programs also include training of graphics developers who design new ways of visualizing information. Universities are offering undergraduate and graduate programs in analytics in all of these disciplines, though they may be labeled differently. A major growth frontier has been certificate programs in analytics to enable current professionals to re-train and re-tool themselves for analytics careers. Certificate programs enable practicing analysts to gain basic proficiency in specific software by taking a few critical courses. Power (2012) published a partial list of the graduate programs in analytics, but there are likely many more such programs, with new ones being added daily [11]. Watson (2013) describes how he teaches his analytics courses and also argues for the growing popularity of such programs [17]. Graduates from these programs fill such positions as data analysts, data scientists and data mining architects.

Another group of players assist in developing competency in the field of analytics. These are certification programs awarding certificates of expertise in specific software. Most (if not all) major technology providers (IBM, Microsoft, Microstrategy, Oracle, SAS, Teradata, etc.) have their own certification program. These certificates ensure that potential new hires have a certain level of tool skills. The Certified Analytics Professional (CAP) certificate program introduced by INFROMS has the main aim of equipping individuals with a general competency in data analytics. Any of these certifications gives a college student additional marketable skills.

Many of these industry leaders also have academic alliance programs to enable teaching their

software and its uses. IBM, Microsoft, Oracle, SAP, and Teradata have major initiatives that allow faculty members to use and share software as well as other teaching materials for instructional purposes. Watson (2013) provides a summary of several of these resources [17]. For example, Teradata University Network (www.teradatauniversitynetwork.com) also known as TUN is sponsored and supported by Teradata, a major provider of analytics hardware/software and applications consulting services. TUN is free to use and provides faculty members and students access to many case studies, white papers, teaching materials, and software for learning concepts, applications, and tools for analytics. Faculty members contribute such resources to TUN. Even teaching notes are available for many of the teaching materials. A recent addition is several videos that depict analytics applications in realistic settings. These videos have been developed by Dr David Schrader of Teradata. These are labeled Business Scenario Investigations (BSI), a takeoff on the popular TV show *CSI*. Not only are these entertaining, but they also provide the class some questions for discussion. For example, in at <http://www.teradatauniversitynetwork.com/teach-and-learn/library-item/?LibraryItemId=889>, one can assume the role of an airline customer service center professional. A video on YouTube gives information about customers' profiles and relationship with the airline. An incoming flight is running late, and several passengers are likely to miss their connecting flights. There are seats on one outgoing flight that can accommodate two of the four passengers. Students watch the video, pause it as appropriate, and answer the questions on which passengers should be given priority. They then resume the video to get more information. Students' decisions might change as they learn more about those customers' profiles. TUN also provides information about how the analysis was prepared in a link at the bottom of the video page. This engaging multimedia excursion provides an example of how additional information available through an enterprise data warehouse can assist in decision making. There are almost a dozen such videos and hundreds of other teaching resources available through TUN.

6. Conclusion

Overall, there is much to be excited about in the analytics industry at this point. Its applications stretch to almost all fields of practice and research including healthcare, retail and manufacturing. Support for teaching analytics is available from many major vendors. The industry clusters of analytics ecosystem show that there are many opportunities for students, academics and industry professionals. With increases in data availability and the realization that more business insights could be generated from an otherwise junk volume of data, research in Big Data has increased. This trend in Big Data analytics is expected to continue to grow and complement traditional data analytics methods and existing data management strategies. The increase in the amount of attention analytics is receiving from the press also indicates that this is an expanding field which can provide a competitive edge to many types of businesses. As illustrated in this paper, descriptive, predictive and prescriptive analytics applications and research opportunities are wide and growing.

References

- [1] N. ARMSTRONG, J. POWEL, Patient perspectives on health advice posted on Internet discussion boards: a qualitative study. *Health Expectations*, **12**(13) (2009), 313–320.
- [2] G. BOND, R. BURR, F. WOLF, M. PRICE, S. MCCURRY, L. TERI, The Effects of a Web-Based Intervention on the Physical Outcomes Associated with Diabetes among Adults Aged 60 and Older: A Randomized Trial. *Diabetes Technology and Therapeutics*, **9**(1) (2007), 52–59.
- [3] C.-L. DENNIS, Peer support for postpartum depression: volunteers' perceptions, recruitment strategies and training from a randomized controlled trial. *Health Promotion International*, **28**(2) (2013), 187–196.
- [4] M. Evans, L. Donelle, L. Hume-Loveland, Social support and online postpartum depression discussion groups: A content analysis. *Patient Education and Counseling*, **87**(3) (2012), 405–410.
- [5] INFORMS, Analytics Section Overview. <http://www.informs.org/Community/Analytics>, Accessed February 2013.
- [6] B. LAKEY, S. COHEN, Social support theory and measurement. In S. COHEN, L. G. UNDERWOOD, B. GOTTLIEB (EDS.), *Social Support Measurement and Intervention: A Guide for Health and Social Scientists*. Oxford University Press, (2000) Toronto.

- [7] M. E. LARA, J. LEADER, D. N. KLEIN, The association between social support and course of depression: Is it confounded with personality? *Journal of Abnormal Psychology*, **106** (1997), 478–482.
- [8] E. K. LEE, F. PIETZ, B. BENECKE, J. MASON, G. BUREL, Advancing Public Health and Medical Preparedness with Operations Research. *Interfaces*, **43**(1) (2013), 79–98.
- [9] M. P. MARTINASEK, A. D. PANZERA, T. SCHNEIDER, J. H. LINDENBERGER, C. A. BRYANT, R. J. MCDERMOTT, M. COULURIS, Benefits and Barriers of Pediatric Healthcare Providers toward Using Social Media in Asthma Care. *American Journal for Health Education*, **42**(4) (2011).
- [10] P. PEKGUN, R. P. MENICH, S. ACHARYA, P. G. FINCH, F. DESCHAMPS, K. MALLERY, J. V. SISTINE, K. CHRISTIANSON, F. J. CARLSON REZIDOR, Hotel Group Maximizes Revenue Through Improved Demand Management and Price Optimization. *Interfaces*, **43**(1) (2013), 21–36.
- [11] D. P. POWER, What universities offer master's degrees in analytics and data science. (2012). <http://dssresources.com/faq/index.php?action=artikel&id=250>, Accessed Feb 2013.
- [12] J. SARASOHN-KAHN, The wisdom of patients: Health care meets online social media. Oakland, CA: California HealthCare Foundation. (2008). <http://www.chcf.org/publications/2008/04/the-wisdom-of-patients-health-care-meets-online-social-media>, Accessed Oct 2012.
- [13] SAS.COM, Eight Levels of Analytics http://www.sas.com/news/sascom/analytics_levels.pdf, Accessed Feb 2013.
- [14] I. SKARSATER, A. LANGIUS, H. AGREN, L. HAGSTROM, K. DENCKER, Sense of coherence and social support in relation to recovery in first-episode patients with major depression: A one-year prospective study. *International Journal of Mental Health Nursing*, **14** (2005), 258–264.
- [15] M. STEWART, V. TILDEN, The contributions of health care science to social support. *International Journal of Nursing Studies*, **32** (2005), 535–544.
- [16] T. VARELAS, S. ARCHONTAKI, J. DIMOTIKALIS, O. TURAN, I. LAZAKIS, O. VARELAS, Optimizing Ship Routing to Maximize Fleet Revenue at Danaos. *Interface*, **43**(1) (2013), 37–47.
- [17] H. WATSON, The Business Case for Analytics. *BizEd*, May/June (2013), 49–54.
- [18] R. SHARDA, D. ASAMOAH, N. PONNA, Business Analytics: Research and Teaching Perspectives, *Proceedings of the 35th International Conference on Information Technology Interfaces*, Dubrovnik, Croatia, June 2013.

Received: August, 2013
Revised: September, 2013
Accepted: September, 2013

Contact addresses:

Ramesh Sharda
Institute for Research in Information Systems
Spears School of Business
Oklahoma State University
Stillwater, OK 74078
USA
e-mail: ramesh.sharda@okstate.edu

Daniel Adomako Asamoah
Institute for Research in Information Systems
Spears School of Business
Oklahoma State University
Stillwater, OK 74078
USA
e-mail: asamoah@okstate.edu

Natraj Ponna
Institute for Research in Information Systems
Spears School of Business
Oklahoma State University
Stillwater, OK 74078
USA
e-mail: natraj.ponna@okstate.edu

RAMESH SHARDA is Director of the PhD in Business for Executives Program, the Institute for Research in Information Systems (IRIS), ConocoPhillips Chair and Regents Professor of management science and information systems in the Spears School of Business at Oklahoma State University. He has co-authored two text books (*Decision Support and Business Intelligence Systems*, 9th edition, Prentice Hall and *Business Intelligence: A Managerial Approach*, 2nd Edition, Prentice Hall). His research has been published in major journals in management science and information systems including *Management Science*, *Operations Research*, *Information Systems Research*, *Decision Support Systems*, *Interfaces*, *INFORMS Journal on Computing*, and many others. He is a member of the editorial boards of journals such as the *Decision Support Systems*, *ACM Transactions of MIS*, and *Information Systems Frontiers*. He is currently serving as the Executive Director of Teradata University Network. He recently received the 2013 INFORMS HG Computing Society Lifetime Service Award.

DANIEL ADOMAKO ASAMOAH is pursuing a PhD in management science and information systems in the Spears School of Business at the Oklahoma State University. He obtained a Master of Science degree in telecommunications management from the same university. Prior to his graduate studies, he completed a Bachelor of Science degree in electrical and electronic engineering at the Kwame Nkrumah University of Science and Technology, Ghana. His research has been published in journals such as *Decision Support Systems and Simulation: Transactions of the Society for Modeling and Simulation International*. His current research interests include the use of Big Data analytics as a platform to study the role of social media in healthcare provision. His research also involves business intelligence, data analytics, decisions support systems in healthcare and the use of analytical modeling in solving operations management and information systems problems.

NATRAJ PONNA is a Master's student in management information systems at Oklahoma State University. He is a BASE SAS®9 certified programmer and a certified SAS®predictive modeler with experience in using SAS Enterprise Miner 7. He also has the SAS®and OSU Data Mining Certificate and has undertaken several projects in business intelligence and data mining.
