**Imperial Journal of Interdisciplinary Research (IJIR)**
**Vol-3, Issue-4, 2017**
**ISSN: 2454-1362,** http://www.onlinejournal.in

# Big Data and Analytics: The Future of Business

Suhel Karara
Narsee Monjee Institute of Management Studies, Mumbai-India

***Abstract:*** *Big data analytics is using large scale data to provide valuable insights into business models. The 4 characteristics of big data include: volume, velocity, variety and veracity. Several industries are using big data analytics for predicting customer behaviors, root cause analysis and preventive treatments. The clinical domain in the Pharmaceutical industry represents a mature analytical industry and has extensively used analytics in optimization of clinical trials and creating adaptive patient centric trials. The sports industry is an excellent example of an exploratory phase of incorporating big data analytics to predict player performance, injury prevention and capitalizing fan preferences.*

## 1. Introduction

Big data is described as large amount of data (structured/unstructured) which are generated by business during their daily operations. This data has invaluable insights for decision making. It is commonly described by the following four V's:

**Volume:**
Big data as the name refers consists of huge volumes of data from different sources like social media, customer attitudinal data, machine data etc. Previously handling such volumes of data was difficult, tools like Hadoop have made it efficient to do so. It is predicted that 40 Zitabytes of data will be created by 2020 which is approximately 300 times data since 2005 [1]. Therefore, now is a great opportunity to harness this data into actionable insights.

**Velocity:**
The data speed at which the data is captured is referred to as the velocity. With regards to big data, the velocity is very high for many applications. For example, the New York Stock exchange can handle 1 TB of trade information data during each trading session [1]. The flow of data is referred to as velocity vector.

**Variety**
The different data feed type can be referred to as variety. The data feeds can be structured, un-structured, open data types, emails, texts, images, audio, video etc Nearly 30 billion pieces of content are shared on Facebook every month [1].

**Veracity**
Veracity signifies the quality of data. Inconsistent data types can contain outliers, abnormalities and can be imprecise. Ensuring data quality is significant in order to specify a high confidence level in the analysis. This can have a significant cost as well, poor data quality cost the US economy around $3.1 trillion a year [1].

## 2. Importance of Big Data and Analytics

The importance of big data is not determined by the volume of data available; it is determined by the analysis that can be performed on these huge datasets. These datasets are assets that can be used for optimizing the following business processes: 1) cost reductions, 2) time reductions, 3) new product development and optimized offerings, and 4) smart decision making [2]. Below are some use cases that can benefit from big data analytics:

- Predicting customer behavior based on attitudinal data
- Determining root cause analysis for failures, defects etc
- Proactive fraudulent behavior detection
- Providing diagnostic and treatment options to patients

## 3. Big Data in Action

In the past few years, Big data applications has grown exponentially. The raw data requires data mining followed by analysis and visualization tools in order to harness the full potential. Large scale companies need to create a mind set shift from a knowing to experimental culture to use these datasets. There are several domains where big data is actively being used:

### 3.1. Life Science-Pharmaceutical Domain

Pharmaceutical companies are leveraging big data application to derive clinical trial insights. Having real-time clinical metrics and dashboards which provides insights on patient enrollment, study

conduct, close-out and reporting is one of the biggest challenges for bio-pharmaceutical R&D industry. With a broad range of study designs, varying data collection methods and time points, efficient data analysis in clinical development has become more important than ever [3]. The development of medicine is a complex process involving huge investment in the R&D process especially the design and conduct of clinical trials. The biopharmaceutical industry spent nearly $10 billion directly in the conduct of clinical trials at the site level across the U.S. in 2013 [4]. Therefore, streamlining these trials is an urgent need to improve clinical cycle time, cost savings and determine novel treatments for the patients in need. The increasing cost and complexity of the clinical trials can be overcome by utilizing data analytics in the following clinical domains: subject enrollment forecasting, study design and planning, visibility of enrollment trends and inability to rescue a setback trial [5]. The pharmaceutical companies are using a bottom-up analytical approach to ensure the optimization of clinical trials. The 3-deminsional bottom up approach focuses on analytics at the study → site → country level. Study level application is designed to enable querying across the portfolio of study data in order to manage workload by understanding the volume of upcoming study milestones [5]. The ability to collect and model site enrollment trends and make decisions to shut down or set up additional sites is key to managing a global development portfolio as study designs become more complex and resources become more scarce [5]. The Country Management Analytics should serve as a summary reporting tool to support site/study level clinical performance insights and operations based on geography. This should be leveraged across clinical study manager, global heads, program and portfolio leads across therapeutic areas [5]. This advanced three dimensional approach has several benefits and can be applied to other industries as well to streamline complex processes. A bottom-up analytical framework such as this one ensure optimization at each level.

Data integration and analysis always requires partnership with IT. This partnership has been observed in the Pharmaceutical industry by adopting agile and scrum methodology. Scrum framework was designed to handle complex software development projects but has found great application in the clinical trial management. With the constantly changing needs of the business in the life science domain, developing multi year data projects can be difficult. Agile development helps to compensate for this. It allows rapid delivery of business benefits by allowing information users and creators to collaborate and find the best means to reach a defined and shared business objective [6]. Rapid prototyping and iterative development methods such as Dynamic systems development method (DSDM) and Scrum can clarify key requirements quickly, identifying risks in time to overcome them [6].
Overall the clinical domain has advanced in the filed with incorporating analytics in several aspects. The following figure represents key areas in clinical analytics:

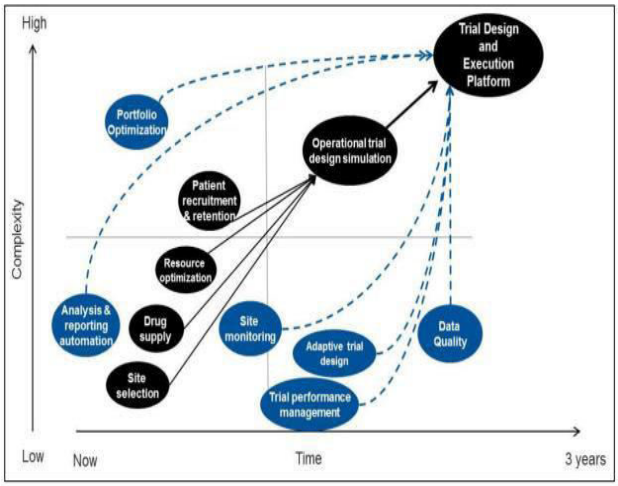

Figure 1. Key Areas in Clinical Analytics [6]

These data points can also be used for calculating, monitoring and reducing risk at various stages in the trial-this concept is known as risk based monitoring (RBM). RBM is the study
of identifying and mitigating the points where a trial and its subjects are at greatest risk for harm through data negligence, lack of education or incorrect conclusions [7]. The intent is to bring intense focus onto the most likely failure points, looking at the data for what is not being revealed during study conduct [7]. There are serval challenges as well in implementing data models and processes, but overall the clinical domain has made significant progress in implementing analytics. The clinical domain can indeed be described as mature in this cycle with several learnings that can benefit other industries as well.

### 3.2. Big Data Analytics in Sports

While the clinical domain has already advanced in data analytics, the sports domain is still in the exploratory phase. There are potential use cases which can provide insightful analysis. In the recent years several pilot programs and wearable technologies have used big data analytics to help players and business owners with data driven decision making. The NFL has rolled out a pilot program with Radio Frequency Identification (RFID) chips which help track the players speed and location. The tags blink 25 times per second and deliver data in 120 milliseconds, thus collecting

billions of datasets resulting in big data analysis. The dashboards and visualizations created on these huge datasets can help coached monitor the player performance like field goal success rate and punt distance distribution [8].

Another great use of data analytics in this domain is injury prevention for the players. Analytics can provide guidance on the player's hydration and physical state. Currently, the response to injuries is reactive. Better insights can make injury prevention more pro-active. There is a huge cost associated with these injuries as well- Major League Baseball spent $665 million in 2013 on the salaries of injured players and their replacements, while the NBA lost $358 million during that same timeframe [9]. Several pilots with wearable technologies have been employed to predict injuries proactively. There was 30% injury reduction in a Kitman labs pilot performed for 3 years with Los Angeles Dodgers and the San Francisco Giants [9]. There has been a surge in the demand for wearable technology in sports which can track several metrics like fatigue rates, health benchmarks and trauma monitoring trends.

Capturing and enhancing sports fan preferences is yet another opportunity for analytics in this segment. Analytical platforms provide a 360-degree fan experience which can be customized based on the attitudinal data. The data elements capture behavioral aspects of the fan engagement, based on the ticket sales, geography, historical games, food preferences etc This can then be leveraged to create predictive data models and what if scenarios. Sports marketing campaigns can target and segment different audience based on these insights. The right channel mix can result in increased ticket sales and revenues.

## 4. Conclusion

Analytics has a huge opportunity to make an impact in several business models. The data streamlines processes reducing cost and increasing revenue. It also helps understand customer preferences to steer the company in the right direction. Data analytics is being extensively used by the clinical domain to reduce cycle time, improve enrollment metrics, design adaptive and patient centric trials. The life science industry has a vison to use analytics to make healthcare more accessible. Data analytics is also found a place in the sports domain and is currently in the exploratory phase. Several pilot programs using RFID and wearable technology can provide data metrics on player's performance as well as help in proactive injury prediction. Another big scale use of analytics to to customize and predict the fan preferences. Cross pollination between different industries using data analytics can help in knowledge sharing and maturation of this domain. Although different industries have specific use cases, the core data methodologies can be used interchangeably.

## 5. References

[1] The Four V's of Big Data. (n.d.). Retrieved April 07, 2017, from http://www.ibmbigdatahub.com/infographic/four-vs-big-data

[2] What is Big Data and why it matters. (n.d.). Retrieved April 07, 2017, from https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

[3] Lodha A (2016) Big data analytics – Clinical integration and visualization. JCTR 2016, Volume 6, Issue 4 ISSN: 2167-0870

[4] Biopharmaceutical Industry-Sponsored Clinical Trials: Impact on State Economies [PDF]. (2015, March). Battelle Technology Partnership Practice.

[5] Lodha A (2016) Clinical Analytics – Transforming Clinical Development through Big Data. IJIR. Vol-2, Issue-10, 2016

[6] Lodha A (2016) Agile: Open Innovation to Revolutionize Pharmaceutical Strategy. IJIR. Vol-2, Issue-12, 2016

[7] Lodha A (2017) Clinical Trial Analytics: Fulfilling the Future of Clinical Operations Insights. IJIR. Vol-3, Issue-1, 2017

[8] Augur, H., The Author Hannah Augur Hannah is a writer, editor and nerd based in Berlin. She's a researcher with knowledge and work in a variety of fields, including 3D printing, fantasy video games and even coffee manufacturing., Douetteau, F., Jeevan, M., Akinfaderin, W., & McNulty, E. (2016, June 17). Professional Sports Get Makeover From Big Data. Retrieved April 07, 2017, from http://dataconomy.com/2016/06/professional-sports-get-makeover-big-data/

[9] Preventing sports injuries with wearable tech and data analysis. (2016, October 11). Retrieved April 07, 2017, from http://bigdata-madesimple.com/preventing-sports-injuries-with-wearable-tech-and-data-analysis/