Chapter 3

PROBLEM FORMULATION AND OBJECTIVES

3.1 General Problems faced by Big Data explained via examples

- The first thing to note is that though big data is very superior at detecting correlations, it never tells us which correlations are meaningful. A big data study might reveal that from 2006 to 2011 the United States murder rate was fit correlated with the market share of Internet Explorer. Both went down sharply. But it’s hard to see there is any elementary relationship between the two. Talking about another example from 1998 to 2007 the number of new cases of autism diagnosed was particularly well interrelated with sales of organic food. Both went up sharply. But on identifying the correlation won’t by itself tell us whether diet has anything to do with autism.

- Second, big data can work well as an addition to scientific inquiry but seldom succeeds as a wholesale replacement. Molecular biologists, would very much like to be able to suppose the three-dimensional collection of proteins from their underlying DNA sequence, and scientists working on the problem use big data as one tool among many. But no scientist thinks you can crack this problem by crunching data alone. No matter how powerful the statistical analysis but one will always necessitate starting with an analysis that relies on an considerate of physics and biochemistry.

- Third, many tools that are based on big data can be simply gamed. For example, big data programs for grading student essays often rely on measures like sentence length and word superiority, which are found to associate well with the scores given by human graders. But once students figure out how such a program works, they start writing long sentences and using unclear words, rather than learning how to actually prepare and write obvious, rational text.
• Fourth, even when the results of a big data analysis aren’t deliberately gamed, they frequently twist out to be less dynamic than they initially seem. In 2009 Google reported that by analyzing flu-related search queries, it had been able to see the extend of the flu as exactly as and more rapidly than the Centers for Disease Control and Prevention. A few years later Google Flu Trends began to get worse and for the last two years it has made more bad predictions than good ones.

• A fifth fright might be called the echo-chamber product, which also stems from the reality that much of big data comes from the web. Whenever the basis of information for a big data analysis is itself a creation of big data, opportunities for nasty cycles abound.

• A sixth worry is the danger of too many correlations. If you look 100 times for correlations between two variables, you find around five fake correlations that appear statistically striking even though there is no definite significant connection between the variables.

• Seventh, big data is prone to give scientifically sound solutions to hopelessly vague questions. For example in the past few months there have been two disconnect attempts to rank people in terms of their historical importance or cultural support. Based on data drawn from Wikipedia, one is the book “Who’s Bigger? Where Historical Figures actually Rank,” by the computer scientist Steven Skiena and the engineer Charles Ward. The other is an M.I.T. Media Lab project called Pantheon. Both efforts get many effects right like Jesus, Lincoln and Shakespeare were surely significant people but both also make some egregious errors. “Who’s Bigger?” claims that Francis Scott Key was the 19th most important poet in history; Pantheon has claimed that Nostradamus was the 20th most important writer in history, well ahead of Jane Austen (78th) and George Eliot (380th). Worse, both projects suggest a false degree of scientific precision with evaluations that are inherently unclear. Big data can reduce anything to a single number, but you shouldn’t be fooled by the appearance of accuracy.

• One last problem of big data is the hype. Champions of big data encourage it as a innovative advance. But even the examples that people give of the successes so it can be accomplished that big data is at its best when analyzing things that are tremendously
common, but often falls short when analyzing things that are less common. For example, programs that use big data to contract with text, such as search engines and translation programs often rely greatly on something called trigrams: sequences of three words in a row (like “in a row”). Reliable statistical information can be compiled about common trigrams, specifically because they emerge frequently. But no existing body of data will yet be large enough to include all the trigrams that people might use, because of the continuing creativity of language.

Big data is here to stay, as it should be. But let’s be practical: It’s an important resource for anyone analyzing data, not a silver bullet.

3.2 Problems faced by Big Data

The size of the data is growing day by day with the exponential increase of the enterprises. For the reason of decision making in an organizations, the call for processing and analysis of huge quantity of data increases. The diverse operations are used for the data dispensation that includes the culling, tagging, highlighting, searching, indexing etc. Data is generated from a lot of sources in the shape of structured as well as unstructured shape. Big data sizes fluctuate from a few dozen terabytes to numerous petabytes of data. The processing and analysis of big quantity of data or producing the precious information is tough task. As the Big data is the most recent technology that can be valuable for the business organizations, so it is necessary that a range of issues and challenges related with this technology should fetch out into light. The two main problems regarding big data are the storage ability and the processing of the data. As big data grows better, so too will its significance in our everyday lives – business and personal. But with great chance comes great challenges, and there is a crowd of upcoming obstacle which we must tackle as we seek to unlock the full possible of analytics.

Adding big data to more conventional medical research is a theatrical shift for scientists. Scientists are used of using data from firmly controlled clinical trials, where they start with a hypothesis and set out to show it with data from their study. With big data, no theory is needed. The data speaks for itself. The shift for researchers doesn’t stop with eliminating the hypothesis. When it comes to medical research, big data is often considered "dirty data,” because it’s collected outside of the standards of a traditional laboratory study. Researchers in health care can
use big data, it just desires to be viewed differently than data from experiments and used in the appropriate context. Using big data in health care also raises moral issues if the data is used without patient permission. Medical records are governed by rules to defend seclusion in ways that customary consumer marketing data is not. At minimum, a big data analytics stage in healthcare must sustain the key functions necessary for processing the data. The criteria for platform assessment may include ease of use, durability, scalability, skill to manage at different levels of granularity, privacy and security enablement, and quality assurance. To thrive, big data analytics in healthcare wants to be packaged so it is menu-driven, user-friendly and clear. Real-time big data analytics is a key requirement in healthcare. The pause between data collection and processing has to be addressed. The vibrant availability of frequent analytics algorithms, models and methods in a pull-down type of menu is also essential for large-scale acceptance. The important supervisory issues of ownership, governance and standards have to be measured. Health care data is seldom uniform, often bumpy, or generated in bequest IT systems with mismatched formats. This great challenge needs to be addressed as well.

Outlined below are the biggest challenges that analytics will run into in the not so distant future.

**Data Silos**

In most enterprises, the data generated by a practical area ends up being the possessions of that group. This leads to two troubles. First, it’s difficult to get a “complete” sight of the data. Consider all the silos and systems that hold data: CRM, ticketing, bug tracking, completion, etc. Getting all the pertinent systems to even talk to each other is a vast challenge. Second, there’s significant cultural discord within organizations. Typically, each group scheming a data silo ends up caring more about their authority and place in a department rather than the success of the organization as a whole. Organizations need to pool their data to discover the answers and get a complete view of their data.

**Data Scientists**

A typical enterprise generally has 10x more IT employees than analysts or data scientists. The process of analysis starts with a line of business demand. IT collects data from a variety of databases and transfers it to data scientists. Large teams of data scientists are deployed who
spend months (or sometimes years) querying the data. Hiring data scientists (with advanced background in statistics, computer science, and some functional expertise) to speed up the process is hard because people with these skills are tremendously limited. The demand and curiosity in data scientists is skyrocketing. What we need is a new class of technologies that amplify the impact data scientists and allow more people to become data scientists.

**Communication**

Finally, the biggest obstacle is to fully understand the potential of data science. Said another way, the logical gap between a data scientist and a business user is so broad that even communicating insights poses a problem. Anything that does not make instinctive sense is often regarded with uncertainty, or not fully understood, by business users, which can direct to missed opportunities. Data scientists and business users need to align, work more closely together, and build faith to solve business problems.

Data is growing and moving more rapidly than healthcare organizations can consume it; 80% of medical data is formless and is clinically relevant. This data resides in multiple places like individual EMRs, lab and imaging systems, physician notes, medical correspondence, claims, CRM systems and finance. Getting access to this precious data and factoring it into clinical and superior analytics is critical to improving care and outcomes, incentivizing the right behavior and driving efficiencies.

Healthcare organizations are leveraging big data technology to arrest all of the information about a patient to get a more absolute view for insight into care coordination and outcomes-based reimbursement models, population health management, and patient engagement and outreach. I opted for this research area for successfully harnessing big data to unleash the potential to achieve the three critical objectives for healthcare transformation:

**Build sustainable healthcare systems**

The healthcare industry is under competitive and legislative pressure to reduce the cost of care, professionally manage resources and improve patient care. Communal changes, such as the way consumers look forward to acquire and receive care; lifestyle choices excluding the use of social media and mobile technology; the sustained rise of chronic disease; and the push to expand
access to main care are transforming the means healthcare is obtained, delivered and paid for. Healthcare organizations are laying the bottom for enterprise health analytics enriched with big data to obtain a more insightful sympathetic of the patient in the condition of which they are and driving effective resource utilization across the healthcare system.

**Collaborate to improve care and outcomes**

Healthcare organizations require improving the superiority and effectiveness of care while refining patient centricity through engagement and healthcare personalization. Regulatory and marketplace changes require a deep understanding and organization of the risks within patient populations in order to compel better outcomes and decrease readmission rates. Understanding the patient in the situation of who they are as individuals is necessary in creating capable programs that make change. This can be achieved with clinical and higher analytics superior with big data.

**Increase access to healthcare**

To advance the health and well-being of our population, we must discover ways to augment access through more effective appointment. An estimated 150M patients in the India in 2010 were repeatedly ill with diseases such as diabetes, congestive heart failure and hypertension, and they accounted for more than 80% of health system costs that year. Attractive and humanizing consumers to make well-versed decisions about defensive care and provider networks can improve health and reduce demand and waste in healthcare.

**3.3 Objectives to be achieved via research work**

Healthcare organizations are leveraging big data technology so that they may arrest all the information about a patient to imagine a total view for insight into care management and outcomes-based reimbursement models, population health management, and patient engagement.

My motivation for this study is to control and explore big data genre to unleash its future perspective in order to achieve these serious objectives for revolutionizing healthcare.
• To construct sustainable healthcare systems which would maintain in the real-world environment despite societal changes?
• To advance care and outcomes collaboratively to minimize readmission rates via minimizing risks management within patient populations.
• To raise access to healthcare to aid people in making informed decisions about precautionary care at initial stages of diseases.
• To persuade and guide people in adopting right living pathway as prevention is always better than cure.
• To provide timely support to people during earlier stages of sickness.
• To give right value of money via reducing scam and mistreatment by practitioners.
• To gain the insight of which symptom expectancy grows among patients above the age of 50 years?
• To know how many teen agers below the age of 20 years are experiencing cardiac linked disorders.

Although not at the scale of some other disciplines, health data sources are big, expanding and complex. It is extensively recognized that they have the power to change medicine, and I am totally strong-minded to play a foremost role in this revolution. My major objective is to make use of big data to full prospective and try my best to bring out conclusions that may timely help to treat dreadful diseases on time before it gets too late.