Data Science for Decision Making in Developing Economies:
A Travesty of Business Investments?

Arindam Banerjee

विद्याविनियोगाद्विकासः

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD-380 015
INDIA

Data Science for Decision Making in Developing Economies:
A Travesty of Business Investments?

Arindam Banerjee
Indian Institute of Management, Ahmedabad
Email: arindam@iimahd.ernet.in

Abstract

Based on a recent survey of Analytics Adoption in Indian business organization, the author makes a claim that for most organizations grappling with the problem of incomplete and unorganized data, the tools of data science are mostly unhelpful in providing impactful information for decision making.

What is required for most organizations embarking upon the Analytics journey is a strategic information plan which assesses requirement, availability and enhancement of information to support key decisions of the organization. This requires the generic skills of problem structuring and information resource identification and mapping them to business decision making. Knowledge of processing tools is important to the extent that information strategists are able to visualize what data and how to process them together to extract the right information.

Therefore the top most priority in organizations in nascent environments is to work towards building a necessary healthy ecosystem of information which in the future will facilitate more analytics driven decision making.

Data Science for Decision Making in Developing Economies:
A Travesty of Business Investments?

Use the generic skills of a data scientist, but defer the use of their tool box for a more opportune time.

We start this article on the premise that, a) Analytics (also called Data Science) is a fairly debated subject matter in the current realm of business affairs in many parts of the world and, b) it appears to be an exciting field for professionals to look for opportunities to build their career in India. Hence, it is presumably a topical issue of importance to many students of management.

However, "Data Science in Developing Economies" appears to be a contradiction of sorts. The fact is that the basic motivation behind the practice of Data science is building efficiencies in a business system in an environment that is well saturated and holds little opportunities for identifying new turfs of opportunity. On the contrary, going by the general definition of a 'Developing Economy", the new opportunity turf in such economies exceeds, by leaps and bounds, the market scape that is currently under the influence of business manipulations. In this scenario, trying to use the virtues of data science and predictive modelling using traditional sources of data is like isolating opportunities from thin files of information and extrapolating the same into a vast ocean of unknown and untrodden space. This is by itself can be an anecdote for ineffective utilization of limited organizational resources.

The traditional scope of data science as we know it is best for extracting valuable information from existing data resources collated from past transactions and is usually worthwhile to decide on allocation of resources appropriately to present business context. The complication in a developing economy is that often times the actual opportunities for successful deployment of organizational resources are nascent and certainly not identifiable from past responses since the subset of responses is far less in comparison to the probable scale of the universe of opportunities.

What does a Data Scientist do?

We interpret IBM[1]'s definition of its data scientists as someone who has a combination of skills in the realm of math/statistics, computer science and data warehousing and is able to combine data of different formats and sources together to provide meaningful insights and inferences that are useful for business decision making. We gathered that the emphasis is more on the extraction of information in meaningful ways utilizing technological skills from associated disciplines to help managers understand complex issues in a more simplistic manner. They are the translators of alien languages stored in varied formats in computing infrastructure into business language. However, the driver of such pursuits is the accessibility to data, something that appears to be scarce in the emerging markets.

---

[1] http://www-01.ibm.com/software/data/infosphere/data-scientist/ (Accessed on December 19, 2014 18:00 hrs (IST))

The latest trend in data surfeit economies seemingly is about the "internet of data"[2] which requires skills sets to mine multiple sets of interconnected data organizations simultaneously and link their insights to form a clearer vision of the future business imperatives for business organizations.

This is perhaps not the suitable model of analytics deployment in a developing economy at the moment. To substantiate this claim, at least partly, it may be worthwhile to look at the uniqueness of the context of India, a representative developing economy, and which appears to have dimensions that may require special attention.

Complications in data sourcing in emerging economies like India:

a) **Non availability of Comprehensive Business Data:** A prerequisite for effective data science application is the availability of data. It may be structured or semi-structured (or unstructured), but the nevertheless it is important that the coverage of the available data source should be close to complete and the variety of information available is broad enough to provide a wholesome view of the business phenomenon that is studied. None of these conditions satisfy in the average developing economy, especially India where information collection architecture is very basic to be able to generate anywhere close to a holistic view of the business environment. In fact, such is the scale of incompleteness that many organizations struggle to initiate systematic planning of their existing information resources.

External information regarding markets and environment are the most difficult to acquire simply because there are few private or government agencies involved in the collection processes. Besides, the high cost of collection of data from relatively inaccessible parts of the country (rural markets, for instance) discourages investments in such initiatives. Even the government of India depends upon the local policing infrastructure to monitor and validate the authenticity of its citizenry while issuing passports. It is hoped that the Aadhaar card initiative will tide over some of these perennial bottleneck s of tracking residents of the country.

A consequent problem due to this no availability of data is that the impact of data and its subsequent processing and insight on decision making remains largely muted and incomplete.

b) **Internal data in multiple and incompatible formats:** A second dimension of the complication for some organizations that have quality business data and that get generated as a part of the business operations, such as transaction data in banks and retail stores, is their availability in different formats which causes immense problems of consolidation. Take for instance the banking and financial services institutions in India. In the past decade and a half, there has been rapid development in computerization and automation of operations in most large public sector institutions. A consequence of this has been that while recent data is available in standardized electronic formats, their integration (or lack of it) with data

---

[2] Frontiers of Computer Science, December 2012, Volume 6, Issue 6, pp 660-667

available in legacy physical systems (read: paper formats) makes it difficult to apply any data science procedures reliably to glean insights for decision making.

c) **Dependency on Heuristics for making decisions:** Given the above constraints, many business organizations in developing countries including India remain steadfast on their dependency on heuristic business rules developed over long periods of experience and a firm connect "with the ground". People –driven decisions override attempts at standardization and the common refrain heard is that information is not available or incomplete to substitute the "gut feel" with the rigors of scientific models based decision support systems". A notable example of such a focus is the role of branch operations in managing business operations in the field. It is very apparent in rural markets, where the role of the local branch is important for taking both operational and at times strategic decisions. Here the lack of information is substituted by the "look and feel "of the environment, which is only possible through a decentralized branch based operations. A centralized process of decision making using data is therefore dispensed with and substituted by a people led decentralized organization structure.

d) **Market Growth hides the virtues of Analytics-driven precision in Decisions:** The futility of the analytics practice is driven by the notion of the "growing market syndrome". Data scientists are supposed to extract business insights that act as a welcome succour in a highly mature and penetrated market. They are supposed to provide directions, refine decisions to hone in on the "close to the perfect" set of decisions for any environment. However, when the markets are in the expanding phase, such extraction of precise insights from past transactions is not quite relevant. In such a situation the importance of factual evidences, based on past occurrences can easily be discounted since the growth in the market overrides the leakages of a sub-optimal decision. Precision in decision making or the lack of it has little consequence on the year end performance of the organization since the overall market growth covers up for all such inefficiencies.

A gist of these complications across various industries in the Indian context (based on a fact finding initiative by the author) are listed Table 1. It is apparent that the role of precision driven model – based decisioning system such as the ones projected by the experts in the realm of data science is far-fetched when compared to the requirements of most industries (except the off shore domain) in a developing economy context.

But then is there a role for traditional analytical processes in decision making in such economies and what are the potential priority areas for investment? In the next section, we shall attempt to answer this question.

So what should Analytical systems look like in emerging economies with nascent data infrastructure?

We will start this section with a conjecture on what may be the pitfalls of a purely subjective / heuristic based decision process in emerging markets. Overtly the risks involved are minimal – so long as a breed of very capable and effective decision makers are available to take the right

decisions. However, this situation by its very nature appears to be serendipitous. There is obviously no guarantee that subjective capability is sustainable in organizations in the long run. Hence, some elements of objectivity in the system in terms of a process or, a plan to build one in the long term is required to act as an insurance against possible scarcity of high calibre leadership in the organization.

The presumption here is that data science related processes, in spite of their limited capabilities due to the constraints cited earlier do have a role to play, either in terms of partial support, or validation of hunches, or in the best case scenario a new revelation that may have been hidden by normal subjectivity.

However, before we delve into the virtues of data science further, even in the realm of sparse data availability, let me point to some possible action steps for organizations in India (developing world) to consider, especially when there are no clear directions for building appropriate analytic ecosystems.

We list out some guidelines that may be useful for analytics leaders in India. These are largely based on our experience of working with organizations wanting to build analytical decision support systems, but are constrained by having unorganized and distinctly incomplete information resources.

1) **Ignore the data scientists for the time being but, imbibe their approach:** Many doctoral students in management (including the author) train themselves to become rigorous management scientist – they work hard on mathematical problems with fleeting relevance to the management world and devised seemingly eclectic solutions for very narrowly defined problems that perhaps had very little to provide in terms of assistance in real life applications. However, in spite of the many naysayers of this kind of training, it does have an impact in creating a knowledge base which over time trickles down to the application domain. Additionally, and the point that needs emphasis here, it provides a very strong foundation for critical thinking and the clarity in identifying the most appropriate dimensions for problem solving. Data science (and more generally, decision science) has a clear approach towards identifying relevant dimensions, their associations and subsequent impact on business performance. The depth in the approach helps provide perceptive thinking ability, which is essential for developing an effective analytical-driven process for decision making in the long run. Hence, data scientists are required in organizations not just for their business model building skills, but to apply their generic logical problem solving skills towards building better information supported decision making process in the long run.

2) **Develop a medium range strategic analytics plan for addressing topical business questions:** The idea here is to use the data scientist's skills to develop medium range plans to address ongoing business challenges with factual information. To initiate, a) identify what may be very desirable information and seek out potential sources of such information and, b) Identify potential analysis at a high level that may be applied to get answers to business questions that require inputs. It is important that such exercise be done comprehensively and the actual paucity of data needed not be used as basis for making a sub optimal plan. The time for making compromises in implementation will come at a later stage.

Organizations must develop a plan for the ideal situation which acts as a template for identifying ongoing opportunities to improve. **See Figure 1**

3) **Identify near term opportunities for supporting decision making with tactical and "piece meal" analysis:** All planning exercises require a "fall back" option to ensure that expectations are met realistically. Obviously, we should not dilute the importance of comprehensive planning – since they are a roadmap for future enhancements. However, realistic expectations in the near term are set based on availability of data and more importantly, the nature of insights that they provide (under specific assumptions) to help in decision making. This planning activity is very necessary to claim organizational support for initiating near term investment in analytical warehouse.

4) **Apply your organization's data science resources judiciously to provide factual support for decision making as identified in the previous step:** We are perhaps over the hill in terms of organizational acceptance of analytics and its role in decision making. The rest is to make it happen and show results and benefits of the process. The motto is to start simple and here the skills of data science are very helpful. More importantly, data scientists are expected to parsimoniously (but effectively) provide insights that are helpful for decision making with necessary caveats for the decision maker to appreciate the true value of the information. Hence, the role of the data scientists is not just to churn the data, but also provide meaningful inferences in the context of the decision that needs to be made. This may also involve applying qualitative assessment and intuition based on past experience of analysis and also conducting experimentation to confirm partially developed hunches. This capability of building business oriented inferences looking a network of information across many and disjointed sources (both factual and biased), is a capability that is often times overlooked in today's technology-led forays in analytics.

5) **Look for opportunities to enhance data warehouse with newer sources of internal and external data that can support better analytics infrastructure:** Last but not the least, for long term sustenance, the process of enhancing business data warehouse is an ongoing one. As markets develop and collection infrastructure improves, the organization must step in to leverage the advantage of newer and better information and augment its current analytical processes. In the process, ensure the fulfilment of a more comprehensive decision support system. Hopefully, that would be the harbinger to enhance the role of data science in business decision making in developing economies.

Concluding Remarks:

This is an exacting set of actions for most organizations. However, with increasing competition and lesser market space to manoeuvre, the sooner organizations tread down this path the better for them to retain their precision in decision making. Surely, the implementation of data science in a direct sense is easier, provided the data are available in a form that readily lends themselves to processing and insight generation. For most organizations in developing economies, there is much work to be done in data management preceding such direct and impactful use of data tools. Barring

some domains such as digital commerce (which are about 5% of the total transaction in India[3]), most business organizations do not have the luxury to an organized access to information. Hence, Analytics processes in countries like India require a different focus compared to what is seen in the developed world.

Table 1 (Summary of Findings across major Industries)
Basis: IIM Ahmedabad interviews with 25 executives associated with Analytics across industries in India (2014)

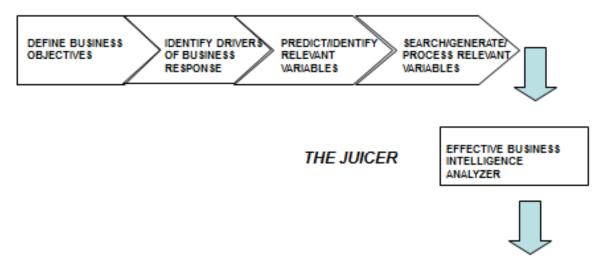| Sector	Key Achievements	Challenges for the future
Large Public Sector banks	a) Electronic Data Capture is mostly complete b) Data warehousing and report generation in progress	a) Do they need sophisticated Analytics to support the banking function in a regulated market? b) Data resources is a mish-mash of electronic and old paper formats. It is very hard to organize them systematically.
Consumer Products / FMCG	a) High competition is compelling management to turn attention on Analytics for better planning b) Better skills available with global movement of human resource	a) Scrappy data management and collection impedes effective output b) Disconnect between Analytics processors and users of its output – don't understand each other well.
Manufacturing / Engineering	a) Effective Analytics capability available for focussed engineering applications b) Requisite skill set is available to support such analysis and inferencing	a) Non availability of comprehensive business databases (especially market related) hampers the true potential of activity.
ITES / Offshore consultants / Captives	a) Developed processes and databases b) Supporting business decisioning with at least baseline support through reporting and analysis	a) Disconnect between Analytical prowess and Business imperatives due to geographic distance b) How to add value through Analytics?
Analytics Consultants for Indian Organizations	a) Provide analytical Services (operational) for direct mailing and customer targeting	a) Not enough scale in the Analytical Services space b) Not enough going on in strategic advisory through

---

[3] http://www.pwc.in/assets/pdfs/publications/2014/evolution-of-e-commerce-in-india.pdf, accessed on December 22, 2014 (15:30 hrs).

Sector	Key Achievements	Challenges for the future
	activities. b) Reporting services for digital and web-based data	Analytics due to scattered data bases in most indigenous organizations c) Some organizations with large databases prefer to invest in an in house consulting / analytical support. d) Forced to look overseas for offshore opportunities.
Financial Services	a) New businesses are introducing relevant data collection architecture with the hope that they can collect relevant market information for future use. b) Some tactical level programs are support with Analytics	a) Felt need is low in organizations for Analytics as the market continues to grow. b) Data exists in various forms – electronic, paper and sometimes on in experience of employees. It is hard to put them together to run a reasonable analytical process for supporting decisioning.

Figure 1

A Smart Analytics Planning Process

THE Process to Develop and Maintain Superior Quality Data

DEFINE BUSINESS OBJECTIVES → IDENTIFY DRIVERS OF BUSINESS RESPONSE → PREDICT/IDENTIFY RELEVANT VARIABLES → SEARCH/GENERATE/PROCESS RELEVANT VARIABLES

THE JUICER EFFECTIVE BUSINESS INTELLIGENCE ANALYZER

THE JUICE....RELEVANT BUSINESS INSIGHTS