The Monetization of Big Data: Made Possible By Humans

The Monetization of Big Data: Made Possible by Humans

Ever since Herman Hollerith patented Electric Tabulating Machines, Big Data has become an essential part of daily life. As necessity is the mother of invention, Hollerith saw a need for machinery to process large sets of information used for processes like the US census. Similarly, a new crop of Big Data companies has been born from companies from all industries integrating Big Data into their daily business operations. Simply put, data is more popular and necessary to companies than ever before. The term “Big Data” has taken on new meaning as it now defines the way algorithms translate petabytes or exabytes of data into actionable insights. Big Data has become an industry unto itself and venture capitalists have taken notice.

The boom of Big Data has been legitimized by the surge of funds invested in data related startups. The category soaked up $2.47 billion in venture capital investments in 2011, a significant jump over 2010's $1.53 billion, according to Thomson Reuters. Venture capital firms like Google Ventures and Greylock Partners have hired in-house data scientists to scout out hot big data companies on the back of the massive IPO of data analytics platform Splunk. The goal of all Big Data companies is simple: monetize data sets. As companies like Target spend millions every year in analyzing consumer data, Big Data companies clamor to aggregate, analyze and provide actionable insights to the companies that will pay top dollar.

As Big Data's popularity soars, startups are beginning to learn that machines can only take them so far. Algorithms can learn, they can predict, and they can recommend. But no matter how complex and powerful they are, they lack intuition and judgement, something only humans can supply. Algorithms and big data get companies 70 percent of the way to their goals. The other 30 percent requires humans. This does not come as welcome news to tech companies. More than any sector, tech pros pride themselves on efficiency and their abilities to solve problems with machines. The last thing they want is to throw expensive human capital at a problem. Often tasks requiring humans are repetitive and, like the big data itself, mammoth in size. To handle the larger-than-life challenges posed by Big Data, businesses are turning to humans, not machines. More specifically, companies are turning to big data solutions provided via outsourcing and crowdsourcing solutions because it is the only solution able to scale, monetize data and process data while avoiding annotative fatigue. Outsourcing and crowdsourcing is not a good option for Big Data companies and projects, it’s the only option.

Scaling: An Adjustable Workforce

The problem with combining humans and big data is one of scale. How do cash-strapped and time-crunched tech companies scale quickly with humans? It would be impossible to pay US wages to complete these tasks in a way that would allow for reasonable ROI. Beyond that, training and hiring armies of data crunchers and curators is prohibitively time-consuming.

Data will remain a continuously growing entity. With over 400 million tweets per day, more than 1 billion photos added to instagram and 995 million monthly active users on Facebook, social data alone is growing at an alarming rate! Social networks like Facebook have created some of

the largest databases of consumer information as they continuously collect, store and organize information created by their global user community. The Internet serves as the world's foremost reserve of data. As with any other marketplaces, the demand for Big Data experiences peaks and descents. For example, as the 2012 Presidential Election neared, Twitter prepared for a sudden increase in demand for social political data. Joining forces with Big Data company Topsy Labs and two reputable polling companies, Twitter analyzed over 400 million tweets a day to determine how people feel about the presidential candidates. Spikes in demand for specific data is common and Big Data companies must organize their infrastructure in such a way that allows them to adapt and react to market needs.

In regards to scalability, crowdsourcing is superior to any other workforce solution. The ability to instantly scale their data categorization workforce by hundreds of employees is incredibly important to companies like Channel Intelligence. Specializing in product data intended to aid manufacturers, retailers and affiliates in selling more inventory, Channel Intelligence is a Big Data company by definition. Despite adding between 50,000 - 100,000 new products per day to their growing database, Channel Intelligence has intermittent data categorization needs. Hence, a crowdsourced solution is ideal for clearing their backlog which can number as many as 73,000 uncategorized products. The work requires virtually no ability to think subjectively and accuracy is not top priority considering their projects with Mechanical Turk yield a 88.3% accuracy rate. For Channel Intelligence, the scalability and on-demand nature of a crowdsourced solution is incredibly helpful in responding to peaks and lulls in database categorization demands.

Monetizing Data: Turning Data From Property Into Profit

Having the unique data set does equal profitability. Optimizing and presenting the data in the manner which provides the most value to customers is what leads to profitability. A young startup called Hoppit found itself facing this exact challenge. Hoppit provides a search engine for restaurants and bars by ambiance. Searching for intangible qualities like ambiance hasn't been done before precisely because of how subjective and intuitive the quality is. Having come from an engineering background, Hoppit founder, Steven Dziedzic was able to create a series of algorithms designed to identify certain keywords and descriptive language which would indicate particular moods or ambiances. Considering the relatively manageable sample size of Hoppit's data set of a few hundred restaurants and bars per city, their internal data scientists were able to successfully create a unique data set. Hoppit had the data, but decisively had no means for creating a valuable product their users. The missing ingredient was high resolution photos. Utilizing in-house employees to seek and acquire thousands of high resolution photos to match their search results would prove too costly and a waste of internal resources. For this particular project, Hoppit retained an outsourced photo acquisition team from TaskUs. Outsourced employees were perfect for this task because of their ability to think subjectively. To build a valuable user experience, Hoppit needed beautiful images of each restaurant that were also indicative of the associated moods and ambiances.

Like Hoppit, FindtheBest, a comparison site started by an ex-Google employee, has embraced human curation as a driving force behind its data engine. The site helps consumers "find

the best" option for them on anything from a puppy to a new car, using a powerful database of hundreds of millions of listings totaling as many as a billion pieces of information. But the database alone was not effective--FindtheBest's engineers learned very early on that without sorting, structuring and vetting the data by humans, its site would not be very useful. The company's 60-person team does much of its human curation in-house, but FindtheBest outsources tasks such as search engine optimization and product tagging to improve the visibility and usefulness of its vast database.

Annotative Fatigue: A Never Ending Struggle Against Productivity's #1 Enemy

Preserving focus for your top in-house talent is just as important as preserving other resources. A+ players within a company contribute to 80% of your company's total revenue, success and sustainability. Robert Munro of Idibon, a startup in beta phase that seeks to provide language translation technology intended for specific use in Big Data, recently commented on his choice to outsource the bulk of Idibon's data sediment analysis needs, "you have to remember that utilizing outsourcing or the crowd is essential as your in-house analysts will inevitably suffer from annotative fatigue. You should preserve their focus for processes that require the highest level of intelligence, analyzation and focus." Sediment analysis, like many other tedious chores associated with Big Data, is a process that should only involve your in-house team at a surface level. The bulk of the grunt work should be reserved for workforce solutions that afford you greater numbers of employees, faster turnaround time and extremely efficient management structures.

Conclusion

Big data is huge--and it's only getting bigger in time. As it penetrates every aspect of technology, companies will grapple with ways to introduce a human element to their number-crunching algorithms. It doesn't matter which percentage of the work needs a human touch--anywhere from five percent to 30% of the work can require people. But companies have realized that without people somewhere along the way, they have nothing but a bunch of numbers and text. Making sense of it all requires humans. Manual tasks are never minor in scale, particularly when big data is involved. Injecting a project with human capital in an affordable, time-sensitive way requires a better solution. This is why tech companies are going big on outsourcing.

Most companies that synthesize big data have in-house data scientists and engineers a-plenty. They might even have a few curators or editors on staff. But they increasingly need an in-between type of individual to complete commodity tasks requiring human judgement. Call them "data refiners" or "data sous chefs." Working behind the scenes, they do grunt work at scale to clean up big data and make it more useful.

Take one of the largest consumer tech companies in the world as an example. Let's call them Company X. Company X was dissatisfied with the search engine on its e-commerce site. After throwing plenty of the world's smartest engineers at the problem, the company realized fixing

the problem required some old-fashioned manual labor. They hired an army of Taskus "data refiners" to check hundreds of keyword searches on its site. The workers repurposed and retagged results until the keyword searches matched the desired outcome. It's a problem the company in question, like any tech company, would have preferred to have solved through automation, but found that humans just did it better.