CHALLENGES IN PROVIDING EFFECTIVE DATA-DRIVEN BUSINESS ENVIRONMENT FOR DATA SCIENTISTS IN ARTIFICIAL INTELLIGENCE TO CREATE ADDED BUSINESS VALUE

Takuya Kudo[1], Kenichi Matsui[2], Shun Yoshida[2],
Michael Munn[1], Ozlem Celik-Tinmaz[1], Michiko Watanabe[3]
[1]Accenture LLP, United States, [2]Accenture Japan, Ltd., Tokyo, Japan,
[3]Keio Research Institute, Keio University, Tokyo, Japan
takuya.kudo@accenture.com

Data scientists, data engineers, and AI researchers are experts in their fields of mathematical algorithms and computer programming, yet they face a significant challenge in bringing actual business value in the optimization process. In discussing the challenges and both educational and business opportunities, this paper refers to an example of close collaboration between organizations such as KDDI, Aizuwakamatsu, Saga and Accenture LLP in establishing the joint initiatives to help those organizations and their ecosystem partners benefit from data-driven decision making and its new business and/or public policy model using a data science method and skills to generate a disruptive market beyond their current operating models.

INTRODUCTION

Data science can have a transformative impact on businesses. In order to achieve that impact, it is important that organizations and the data scientists both take responsibility to improve the business environment and individual's skills towards that goal. Organizations must provide data scientists with an ever-evolving, fit-for-purpose data science workbench with unlimited access to their data and all required (old and new) technologies and tools for developing data science applications. Organizations must be willing to change business processes quickly to make sure they generate, collect and store the necessary quality and objective data for data scientists to mine.

On the other hand, data scientists should possess sufficient domain knowledge about the business they are working, in addition to the mathematical and technical skills they use on their research. The domain knowledge as such is crucial to be able to translate the insights derived from the data and analytical models into actionable business and policy insights to create new business models and added value to the business.

However, there has been a surge in challenges and issues regarding how to equip future data scientists with such supporting and agile business environment complete with quality data and necessary tools, and at the same time grow their domain knowledge to capture and convert full potential of the data results they produce into added business value. This requires multiple levels of effort on both sides, and failure to develop these requirements will lead to lost opportunities to grow by failing to utilize full power of data science that can potentially have a positive disruptive impact on the current market and create new business models. The key success factor here is how to balance theory and practice.

"Trial and Error" is the key to disruptive ideas because the abilities to imagine, experiment, and learn are critical to driving innovation. Achieving this requires a high-velocity environment. However, all too often, data scientists are not equipped to move fast enough, either because they are saddled with aging technology or the tools and data available to them are outdated or insufficient.

DEMAND FOR LITERACY FOR A DATA ANALYTICS PROJECT MEMBER AT AN ENTERPRISE

When starting new data analysis, it is common to perform a Proof of Concept (PoC). We will first attempt a small project to determine whether the combination of data we already have or will acquire, and analytics techniques supports the hypothesis in this PoC. The main reason for adopting the PoC is based on the cost issue related to retrieving the data. Since we cannot determine the prospects from the results unless we investigate the data, we start with a small step. In a data analysis project in business, it is also common practice for companies holding data to proceed with a data analytics service provider such as Accenture. To make such a collaborative

data analytics PoC project successful, it is possible not only for an analytical service provider but also a company ordering data analytics to obtain better results by being conscious of the following three items.

Sharing Domain Knowledge

When initiating a data analytics project, understanding the domain that generates the data is very important not only from the viewpoint of improving model performance but also for steering the analytics output in an appropriate direction. Even if we aim to improve accuracy without domain understanding, there have been many cases in which the obtained analysis output did not fit into the actual operation and could not be applied. It is important to establish an appropriate KPI that can be applied to the actual business in advance and to build an analytics model for this. In addition, it is necessary to introduce domain knowledge to improve the prediction accuracy of analytics models. For example, in the case of adopting a Bayesian hierarchical model, it is necessary to define the dependency of the statistical model between random variables, but no matter how skilled the analysts are, it is difficult to construct a proper model without understanding the relationship between variables in the real world. In some cases, a competition is held with masking data to measure the analytics practitioner's skill. However, after a certain assessment is completed, it is important to check the feasibility level of its holistic solution. This should be done by disclosing the meanings of the data to design and optimize the outcome of their actual business processes. It is simply because the purpose of data science is not only about math and algorithms but also how to maximize the real business outcomes from data insight.

Understanding the scope

Some parts of machine learning techniques can be regarded as pattern recognition, and statistics (especially inferential statistics) is a framework for estimating a population from the obtained sample data set. Since these are inductive approaches, they are basically techniques for interpolation. The prediction for values outside the range of the obtained data set is called extrapolation. By understanding the concept of this interpolation and extrapolation, we can gain a sense of what is and is not possible in analytics. Big data is often accompanied by high dimensionality and as such it is hard for humans to read and understand those data, but if the pattern they are seeking is embedded in the data, it can be extracted with analytic methods and we can predict the event from the data. It is important to recognize that this type of interpolation problem is an effective target with analytics. For example, predicting sales on rainy days with only sales data from sunny days is an extrapolation problem, because the data acquired is biased and rainy days are an unknown experience. Such extrapolation predictions are often unsuccessful. If you can determine the problem, you can convert the problem to an interpolation by re-acquiring the data on rainy days and changing it to an available interpolation analysis. As such, the validity or effectiveness of any given model critically hinges on the assumption that the distribution of the input and output data are similar.

Acquiring proper data

As mentioned in the second item, to make the problem an interpolation, it is very important to obtain appropriate data. It is not unusual to be asked "How much data is enough for the analysis?" However, except for simple inferential statistics problems, it is difficult to answer this question. There are two perspectives on the concept of data quantities. One is simply the amount of data itself; the other is having qualitatively sufficient data such that it covers the scope we want to predict. Regarding the amount of data, from the statistical viewpoint, this can be judged by determining whether the confidence interval is practically sufficiently narrow. In the case of using machine learning techniques, it is possible to make a certain judgment by determining whether the accuracy score converges on the learning curve. However, with regard to the problem of the qualitatively sufficient number, this cannot be judged only by data, and it is necessary to verify the domain knowledge to which it applies. As in the previous example, it is important to specify what type of data is insufficient and to cover a sufficient range of data based upon the domain knowledge.

SELECT CASE STUDIES
        In the previous chapter, we indicated important points in collaboration. This chapter introduces three specific case studies which further highlight these points in action. First, we examine an example of an industry-university partnership in which Accenture takes advantage of its experience to bring local government with students. Second, the powerful data science is now helping doctors, emergency responders and Saga Prefecture officers work in unison to refine the emergency dispatch process. Third, we look at a deep collaboration of Accenture with KDDI and the development of ARISE Analytics.

CASE STUDY: THE UNIVERSITY OF AIZU
Background and Business Challenge
        Since 2011, Accenture has collaborated with Aizuwakamatsu City, Fukushima Prefecture, and the University of Aizu, one of Japan's largest computer science universities, for the revitalization of Fukushima. Aizuwakamatsu City announced that they would jointly launch a plan to create employment and establish the Accenture Center for Innovation in Fukushima.
        In Aizuwakamatsu City, many people were evacuated from the disaster area of the Great East Japan Earthquake of March 2011, so it was important to create employment and promptly promote industry. On the other hand, the foundation for industry and town planning utilizing IT for new employment in this area was already in place, since Aizuwakamatsu City was focusing on the smart city utilization of renewable energy, such as mega solar power, and the University of Aizu is one of the top educational institutions in the IT education and research field. We took advantage of this foundation and led the development of industries that will play a role in the next generation using statistical education and advanced digital technology research to gain benefits from its unlimited economics of scale without physical location constraints rather than creating jobs at call centers or factories.

Achievements and Lessons Learned
        Since the following year, Accenture has been in charge of lectures on statistical analysis at the University of Aizu where students learn how to practice analytics using open data from the fundamentals of the analysis process below.
1. Understanding a series of analytics processes
2. Practicing the analytics process utilizing open data
3. Experiencing practical analytical skills using statistical tools (e.g. SAS, R, Python) used in actual businesses
4. Understanding of logical thinking and presentation skills and acquisition through practice

        As in the above process, students can gain experience not only through classroom lectures, but also projects adapted to business through exercises from problem definitions, hypothesis planning, data searching, analysis, suggestions, and presentations simultaneously.
        Ultimately, the students at the University of Aizu were able to analyze open data sources to find areas prone to accidents and, as a result, made suggestions to the City Hall of Aizuwakamatsu City. Specifically, the students showed the points where accidents are likely to occur by analyzing the position and acceleration data collected by sensors attached to the city's public cars in combination with the position data of the accident locations disclosed by the police. After these results were acknowledged, we announced the expansion of the Accenture Center for Innovation in Fukushima last year. In addition, as a model of regional creation, the Ministry of Internal Affairs and Communications (MIC) and the Ministry of Economy, Trade, and Industry (METI) pledged to provide continued support.

CASE STUDY: SAGA PREFECTURE EMERGENCY DISPATCH
Background and Business Challenge
        In Japan, the majority of rural municipalities are facing the same challenges, including tight budgets, and stretched health care costs and resources due to the surge of an aging population. In this circumstance, emergency patients can be turned away and lose seconds waiting for first responders to identify a hospital that has the right capacity or expertise to help. The Governor of Saga

Prefecture challenged us to find a better way to optimize the end-to-end emergency dispatch process related to life itself.

Our mission was to shorten the emergency transport time from emergency call to arrival at a hospital, relying on a highly complex data supply chain. By even shaving a minute of transport time, we may be able to save lives.

Achievements

We've conducted for the Saga Prefecture, under the aegis of the prefecture governor, a project dealing with the policy planning process, utilizing a data-driven approach. We analyzed transport data collected from iPads installed inside all emergency vehicles in Saga prefecture, thus enabling us to visualize how patients were being transported and to study how transportation times might be shortened. We discovered that it would be possible to reduce around 40% of the current transportation time, bringing about an average time reduction of 1.3 minutes based upon our robust machine learning algorithm. Our unique applied intelligence technology leveraging open source software and our scientific machine learning approach helped improve the coordination among the government, hospitals and emergency agencies and optimized their emergency dispatch process by utilizing their domain knowledge.

Lessons Learned

In this project, we successfully collaborated with experts who have diverse skills connecting through AWS cloud servers across the globe, leveraging the best talent from the US and Japan.

Furthermore, we focused on people as the key factor of continuing the process, thus educating more than 100 government staffs to utilize data with their hands-on analytics skills. This resulted in the development of citizen-centric services in Saga Prefecture. Lessons learned include:

The power of liquid workforce and cloud servers: We are well connected through cloud servers to bring the best optimized outcomes with worldwide best talented people in addition to local professionals with their domain knowledge and its accurate data.

Artificial Intelligence + People with domain knowledge brings better results: Artificial Intelligence (AI) has a great power in predicting something. But when the human becomes engaged through the accurate data, AI becomes much more powerful to bring better result.

Accenture's innovative practice, Data Science Center of Excellence within Applied Intelligence, provided a pioneering the path on which Japan and other countries in the world, can follow in the future. We believe that this is going to be an important underpinning for any data-driven organization's future growth.

CASE STUDY: ARISE ANALYTICS
Background and Business Challenge

KDDI is the second-largest telecommunications provider in Japan and provides an example of an ideal data science infrastructure. The company needed to address the saturated mobile marketplace and their response was to transform from a traditional telecommunications company into what they dubbed a "life design company." In essence, this means KDDI will harness data from customer touchpoints to provide personalized offers and services relevant to its consumers' behaviors and life contexts.

Achievements

With this vision to transform their business, KDDI partnered with Accenture to establish ARISE Analytics. ARISE's mission was to use artificial intelligence (AI) and data science to improve customer experience value and promote new partner businesses with a major focus on customer perspectives and innovation.

To complete this mission, KDDI encouraged a start-up culture, equipping the program team with the right data and a technology foundation to enable it to focus on game-changing applications. The team had access to real-time data at the most detailed level of granularity from

sensors, such as the data communication modules (DCMs) in connected cars. ARISE's technology landscape has been set up to provide a productive data science workbench. Its features include:

Artificial Intelligence "Single-Brain Engine": A real-time recommendation engine powered by centralized customer data from across KDDI affiliates. It focuses on providing insights and personalized recommendations, value-added services, and advertising through preferred customer channels.

Execution by chatbots: A next-generation AI chat service that provides an innovative customer experience driving higher conversion rates and reducing human-intensive processes.

Real-time, comprehensive data: Developing KDDI's IoT business through a comprehensive picture of the customer genome in real time – and at the speed of business – rather than relying on traditional quad-play data.

This environment gave KDDI's data scientists what they needed to iterate at speed (including data preparation, feature detection, and algorithm development), allowing them to focus on business value realization and innovative customer experience design.

> *"The competitive advantages provided by AI and machine learning with AWS components such as EC2, S3, EMR, RDS, DynamoDB, AWS Lambda, Amazon Kinesis etc. are rapidly generating insights can be turned into business value. This is enabled by flexible access to the large-scale computing power provided by AWS required to train the models. This is how our system is accelerating KDDI's growth." (ARISE Analytics Data Scientist)*

Lessons Learned

In contrast, most companies today look very different from KDDI. Data scientists tend to be provided with limited disintegrated desktop tools, with little ability to experiment. Nevertheless, they are still expected to turn raw data into business-making magic.

They are unable to deliver the alchemy required due to a number of constraints. These include a lack of integration with the business, working with inferior data that is unavailable at the speed and granularity required to make a business impact (read Untrap Data), siloed and limited tools, archaic batch deployment models, and long wait times for data access.

There is little surprise, therefore, that data science is often seen as a cash drain and data scientists become disillusioned and disengaged and eventually leave the organization. There are typically a number of issues underlying the disappointing picture painted above. Among them are

Limited assortment of tools and systems: With significant time spent on data preparation, the little time data scientists are afforded for true value creation is limited by statistical desktop and data-mining tools. Often, the time crunch and limited technology ability to experiment with advanced techniques lead to most data scientists working on BI and reporting tools.

The speed of technology dictates the pace: The data science workbenches at organizations are not evolving at the same speed as technology advances in the industry; therefore, the data science landscape at companies is always playing catch-up. This compromises the data science program and leads to limited use cases the business can put forward.

Lack of industrialized intelligence: Most data science programs fail to reach their full potential because of the complexities involved in promoting enterprise adoption and scale. For instance, edge devices involving thousands of real-time deployments require a great deal of work to keep data models updated and fresh. Without robust model management and a real-time deployment environment, business and data science programs often become disjointed and irrelevant to business.

RECOMMENDATIONS

The gap between aspiration and reality can be bridged. In addition, while this may *seem* a daunting prospect, there are actionable, manageable steps that all companies can take to get started:

First, the data science playground. Maximizing data scientists' effectiveness requires a workbench of tools and systems to help with data preparation [and model deployment?] so that data

scientists can concentrate their time and efforts on the math behind business issues. This means enabling data scientists to use an array of ever-evolving, fit-for-purpose technologies, including AI, analytics programming, integrated development environments (IDEs), machine learning, and content analytics. It is also essential to create a role within the organization that will act as a conduit between business requirements and data science technology needs.

Second, focusing on the business adoption of data science outcomes and model management to keep solutions relevant. Models should be deployed in real time and close to the "edge" to align with the pace of business and managed through an automated ecosystem. To operationalize data science outcomes, continuous model management should be used to keep models fresh and design-led applications should be developed to make them as relevant and accessible to the broader business as possible. Using advanced deep reinforcement learning algorithms that reward optimized behavior will help data science systems remain relevant for longer through automated continuous improvement and free up data scientists' time so that they can focus on key business issues.

Finally, remembering that investment in knowledge returns the highest interest. Overall, the landscape may appear complex and expensive to navigate, but many foundational elements of a data science program can provide enormous benefits without massive outlay. Investing in cloud-based solutions frequently offers flexible commercials and requires a relatively low financial commitment. Deploying open-source technologies can prevent having to pay substantial sums for each new component. Additionally, by focusing on "human" interfaces and business processes rather than a plethora of dashboards, data science outcomes will gain faster traction and broader acceptance across the business.

CONCLUSION

This paper proposed an approach to find the connecting touchpoints between academic research resources and the practical reality to bring real business value. In this sense, it is quite important to utilize data-intensive business environments or public-sector workplaces to proactively use these fields as a strategic talent platform to guide them through human interactions and bring technical expertise to actual business value. To achieve this, leaders in this space should be agnostic regarding technologies and free from internal politics or legislative obligations to avoid constraining themselves. Nevertheless, objective joint ventures or external organizations might play a key role in bringing their domain knowledge with an unclouded perspective and advanced technical expertise. In addition to talent platform scarcities, any future data-driven organization needs to overcome three key constraints to achieve successful iterations by preparing computing resources, a massive amount of dormant data, and hard math. Moreover, computer science-balanced skills can be fully utilized through free open-source software and not limited by software or hardware constraints to unleash the full potential of data science. The new direction and format of KDDI and ARISE Analytics in aligning with Accenture and Academic Alliance may be one of the future states of this data-driven organization for addressing the aforementioned issues. Our approach is still in the pilot phase, but we will keep moving forward while iterating trials and errors through data-intensive decision science.

REFERENCES

Accenture Analytics., Kudo, T., & Hoshina, G. (2016). *Practical Course of Data Analytics.* Tokyo: SHOEISHA.

Kanda, K. (2013). *Accenture to Design Strategy for University of Aizu Revitalization Center*. Tokyo: Accenture LLP.

Accenture (2017). *Helping save lives with analytics*. United States: Accenture LLP.

Kudo, T. (2013). *Introductory book for data analytics starter.* Tokyo: PHP Interface.

Kudo, T., & Hoshina, G. (2013). *Introduction to Data Science.* Tokyo: Nikkei BP.