CHALLENGES AND ISSUES IN DEVELOPING REAL-WORLD CURRICULUM FOR DATA SCIENTISTS IN JAPAN

Takuya Kudo[1], Yoshihiro Hayashi[1], Michiko Watanabe[2],
Tomiaki Morikawa[2], Tomoyuki Furutani[2] and Manabu Iwasaki[3]
[1]Accenture Japan, Ltd., Tokyo, Japan,
[2]Keio Research Institute, Keio University, Tokyo, Japan
[3]Seikei University, Tokyo, Japan
takuya.kudo@accenture.com

Like many countries, Japan has challenges in producing sufficient and appropriately-trained statisticians. In response to recent Big Data demands in society, new courses related to data science are being developed in collaboration with IT companies, and an Institute of Professional Data Scientists established, collaborating with universities to develop a competency framework for professional data scientists. This paper considers a number of challenges, including reducing the domestic shortage of skilled analytics professionals, preparing such graduates for complicated real-world business problems, and developing advanced approaches to collecting and analyzing data that can drive value for the country's businesses and public sector organizations. In discussing the challenges and opportunities, the paper refers to an example of close collaboration between the Keio University Research Institute and Accenture Japan, Ltd, in establishing the "Data Business Creation Laboratory," an initiative to help organizations in Japan benefit from data-driven decision-making using analytical methods and skills.
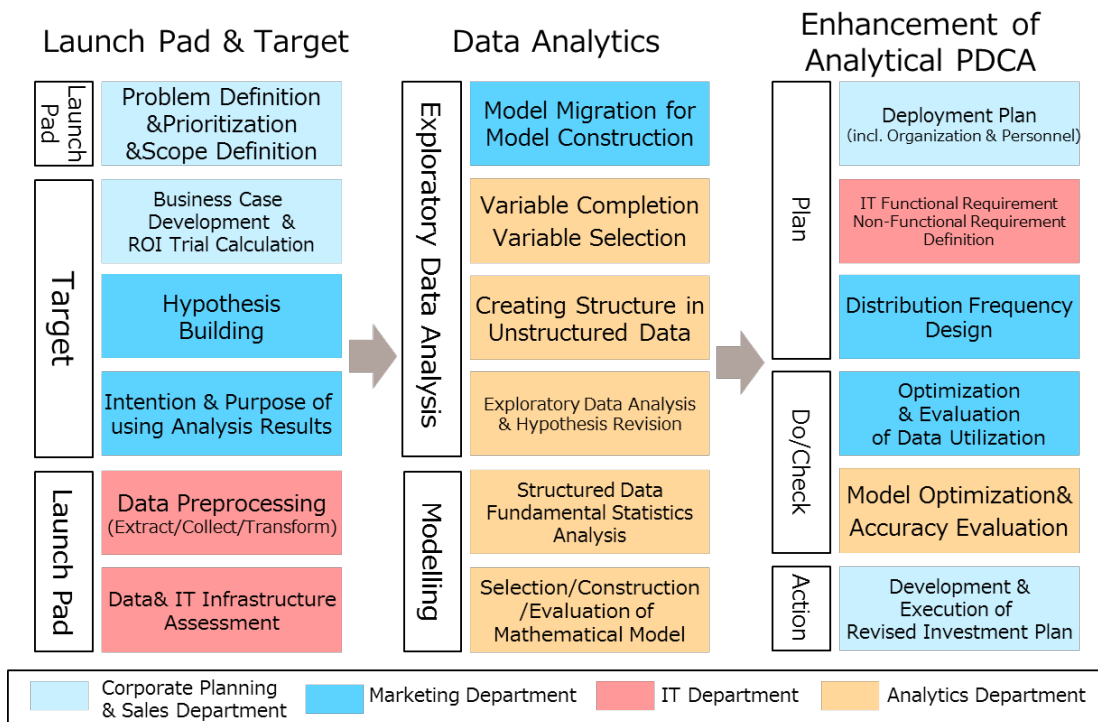
BACKGROUND
With the development of data analytics in Japanese enterprises and the disclosure of public data (open data) by "Abenomics", expectations are rapidly growing for the creation of new businesses using data and the application of new technologies. In response to this movement, the need for human resources who have data analytics skills and can suggest solutions to business challenges has been increasing. On the other hand, in the Japanese academic institutions such as universities which are a major source of human resources for Japanese enterprises and public organizations, Departments of Statistics have not existed in Japan other than The Institute of Statistical Mathematics (as of March 2014) and Statistics has been just one of elective courses in a curriculum in departments such as Medicine, Mathematics, Economics, Psychology, and Engineering, to name just some.

CONTENTS OF PROGRAM
With the aim of launching an institute to foster highly-skilled data scientists and an academia-government collaborative scheme, the Data Management Creation Laboratory was established in the Keio Research Institute at the Shonan Fujiasawa Campus in April 2013. Accenture Japan has been involved with this laboratory since its establishment. Accenture has an "Accenture Analytics Practice" which is a global practice of analytical PDCA (Plan-Do-Check-Action) management cycle consisting of operation design and performance measurement including modeling and machine-learning in several industries, and interpretations of data product. With this global practice, Accenture has introduced many projects of a strategic nature and actively promotes academia-government collaborative projects in several countries, including a collaborative research scheme in North America with MIT (Massachusetts Institute of Technology). The laboratory started on a trial basis in September 2013 by expanding the existing statistics course as an elective liberal arts course into a course integrating Statistics, Computer Science, and Organization Strategies both for public administration and business.
In this program, the contents are organized considering the process of defining business goals and challenges, building and testing hypothesis, and developing prediction models. By providing explanations of statistical analysis methods and how to operate Open Source Software "Revolution R" in an appropriate timing within the course, the methods which Accenture practices are explained systematically. The chart below as found in *Introduction to Data Science* (Kudo & Hoshina, 2013) shows the summary of the analytical process framework starting from current state

analysis and goal setting, which Accenture Japan practices in the case of analyzing marketing data. This chart illustrates the importance of teaching the whole picture of the method of analysis. In Japan, standardized methods have not yet been established and thereby many data analysis projects have ended in failure. This is largely due to projects that are undertaken just to "analyze data" without setting a specific goal. In addition, Japanese enterprises and public organizations are usually organized by functions and each section tends to focus on achieving its own goals without sufficient collaboration. However, the collaboration and support with different sections are essential to succeed in data analysis projects, as is illustrated in this chart which is incorporated into the curriculum.

Launch Pad & Target

Launch Pad	Problem Definition &Prioritization &Scope Definition
Target	Business Case Development & ROI Trial Calculation
	Hypothesis Building
	Intention & Purpose of using Analysis Results
Launch Pad	Data Preprocessing (Extract/Collect/Transform)
	Data& IT Infrastructure Assessment

Data Analytics

Exploratory Data Analysis	Model Migration for Model Construction
	Variable Completion Variable Selection
	Creating Structure in Unstructured Data
	Exploratory Data Analysis & Hypothesis Revision
Modelling	Structured Data Fundamental Statistics Analysis
	Selection/Construction /Evaluation of Mathematical Model

Enhancement of Analytical PDCA

Plan	Deployment Plan (incl. Organization & Personnel)
	IT Functional Requirement Non-Functional Requirement Definition
	Distribution Frequency Design
Do/Check	Optimization & Evaluation of Data Utilization
	Model Optimization& Accuracy Evaluation
Action	Development & Execution of Revised Investment Plan

Corporate Planning & Sales Department	Marketing Department	IT Department	Analytics Department

Source : *Introduction to Data Science*(Kudo & Hoshina, 2013)

In this program, we demonstrate to students that it is important to define "the target" and "the launch pads" first before starting to analyze data. For example, we used the chart above so that students can consider "the target" and "the launch pads" in the lectures about the process of data collection/transformation or data analysis. These help students to understand the relationship between data analysis and business/policy challenges. The program consists of lectures about analytical techniques, and seminars to analyze sample data using "R." The main contents of the program are as follows:

- Business Strategies
- Population and Sample
- Descriptive Statistics
- Probability Distribution(including seminars using "R")
- Hypothesis Testing(including seminars using "R")
- Variance Analysis(including seminars using "R")
- Correlation Analysis(including seminars using "R")

The program consists of lecture materials written by Accenture and textbooks such as *Introduction to Data Science* (Kudo & Hoshina, 2013), *Akaike Information Criterion AIC – Modeling, Prediction and Knowledge Discovery* (Kabashima et al., 2007), *Exploratory Data Analysis* (Tukey, 1997) and *Introductory book for data analytics starter* (Kudo, 2013) as a

reference, and of seminars using "R." Textbooks and case studies are selected from the topics of real business and daily life so that students without business experience can work on them with curiosity and can develop their self-study easily. For example, one lecture used items in a convenience store where students frequently purchase products as data objects in their real life, and analyzed the data to discover that the inconvenience they encounter in their daily life and its relationship to POS (Point-of-Sales) data can be a hint to improve business strategies.

Since this program was given as an elective course in liberal arts and is interdisciplinary, covering different academic areas, some students did not have a basic knowledge about mathematics, statistics and computer science and each student had a different level of computer literacy. To resolve this variety of knowledge and skills of students, graduate students with experience in programming in the department of health management in Keio University together with several consultants from Accenture, cooperated to support students as Teaching Assistants. This succeeded in minimizing the variation in progress among the students, increasing overall performance and improving the level of understanding.

Students varied from freshmen to seniors in college, and to first and second year students in master programs and the course was at capacity with 250 students. Students formed groups of three members from the beginning of the program in September 2013. Through this group work, students learned the diversities in the process of data analysis, and the importance of team collaborations. Student also learned the method of problem solving including defining problems and building hypotheses through producing an interim report in December 2013, preparing for the final report in January 2014. Examples of titles of final reports from students are as follows:

- Analysis on methods to eliminate errors for quality control in a confectionary production process
- Analysis on weather predictions
- Uncertainty pertaining to Human Interactions
- Examination of measures to attract enterprises in the area of Shin-Hakodate Station by analyzing Regional Economies
- Investigation of danger residing in relocations to higher grounds in the tsunami disaster areas
- Analysis on tendencies of successful full-length movies

ACHIEVEMENTS

As an ambitious action to integrating Statistics, Computer Science, and Business Strategies which had not previously happened in Japan, Ichita Yamamoto, Minister of State for Science and Technology Policy, who is in charge of science technology policies and unifies IT policies in Japan, expressed his opinion in an interview with Nikkei Business Online in December 2013, that he would like to promote Keio University's program to foster data human resources of data analytics. In addition, Nikkei BP, a subsidiary company of Nikkei Inc., a prestigious news company which makes "NIKKEI225", a stock index of Tokyo Stock Exchange, reported Keio's program's aim and activity on its web media "IT Pro" in December 2013(2013) As a result, the program gained widespread attention and a solid reputation which led to the launch of a new specialized course in April 2014. We believe that the trial program achieved its initial goal.

FUTURE PLAN

With the aim to foster "human resources who have skills in data analytics and can suggest solutions to business challenges", we will continue to enhance the curriculum. Especially, under significant changes of curriculum across departments in Keio University, a Data Science course has been launched since April 2014. This course is regarded as a specialized course and is titled "Creation of Data Science". In contrast to the trial course which was given as an elective course in liberal arts from September 2013 to January 2014, this new course is limited to about 40 students with more specialized contents, including the following:

- Discussion on location information through analysis of Wi-fi access data
- Discussion on ROI of Data Analysis

We expect that the students from this program will play leading roles in a data analytics domain in Japan, create new technologies and services, and make public policies based on data analysis in the future. As for the contents, we will also develop a topic about machine-learning in addition to descriptive statistics, multivariate statistics, and data-mining.

REFERENCES

Kabashima, Y., Kitagawa, G., Amari, S., Akaike, K., & Shimohira, H. (2007). *Akaike Information Criterion AIC –Modeling, Prediction and Knowledge Discovery*. Tokyo: Kyoritsu Publishing.

Kawamata, H. (Ed.) (2013). *Data Scientist gives morning lectures in Keio University to develop future data analysts in Japan*. Tokyo: IT Pro.

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

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

Tukey, J. W. (1977). *Exploratory Data Analysis*. Boston: PEARSON.