A Study of Practical Education Program on AI, Big Data, and Cloud Computing through Development of Automatic Ordering System

Sachio Saiki*, Naoki Fukuyasu†, Kohei Ichikawa‡, Tetsuya Kanda§,
Masahide Nakamura*, Shinsuke Matsumoto§, Shinichi Yoshida¶, Shinji Kusumoto§

*Graduate School of System Informatics, Kobe University
Email: sachio@carp.kobe-u.ac.jp
†Faculty of Systems Engineering, Wakayama University
‡Graduate School of Science and Technology, Nara Institute of Science and Technology
§Graduate School of Information Science and Technology, Osaka University
¶School of Information, Kochi University of Technology

***Abstract*—An innovative engineer who can address a social challenge using big data processing, AI and cloud computing technologies with the generation of new business and value is required from industry. enPiT is an education project to develop the advanced IT engineer based on practical education in cooperation between industry and academia promoted by Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. In this paper, we introduce how we designed PBL centered curriculum named AiBiC Spiral under the framework of enPiT education project, and analyze educational effect of our program based on work products and the questionnaire result targeting students who took and completed to the program of 2017.**

I. Introduction

With the development of **cloud computing**[1] technologies, various and large-scale information is stored as **big data**[2]. Along with the growth of the big data field, application of **AI** technologies which toward to create various value added are rapidly growing. In the light of social background, innovative engineer who can address a social challenge using big data processing, AI and cloud computing technologies with the generation of new business and value is required from industry. With this requirement, human resource development which can equip students with not only a skill but also social basic ability is also desired. **enPiT AiBiC**[3] is one of a kind education program that aims to develop an innovative systems engineer through the Project Based Learning (PBL) practice mainly targeted at third or forth year undergraduate students from depertment of Computer Science or information Engineering. Therefore, we have started PBL centered curriculum named AiBiC Spiral, which use automatic ordering problem in retail stores as a base for exercise, on Kansai-area of Japan as part of a comprehensive approach to enPiT AiBiC. In this paper, we introduce how we designed a curriculum such as basic knowledge learning and PBL exercises under the framework of enPiT education project, and analyze educational effectiveness of the curriculum based on work products and the questionnaire result targeting students who took and completed to the program of 2017.

II. enPiT AiBiC

enPiT is an educational project to develop the advanced IT engineer based on practical education in cooperation between industry and academia promoted by Ministry of Education, Culture, Sports, Science and Technology (MEXT). enPiT aims to develop the practical engineer of four fields, big data and AI, security, embedded system and business system design based on an education framework. enPiT AiBiC is a brand name of big data, AI, and cloud computing filed in enPiT. enPiT AiBiC is composed of a combination of three hubs divided by area in Japan, **AiBiC Eastern Japan**, **AiBiC Kansai**, and **AiBiC Kyushu** respectively. In each hub, AiBiC program provides practical advanced education collaborate with vendor and user companies of Big data, AI, and Cloud computing fields to undergraduate and college of technology students. In AiBiC Kansai, we accept students from 11 universities and 1 college in Kansai-area, and collaborate number of 21 companies in 2017. A curriculum of AiBiC Kansai is made up of 3 courses, fundamental knowledge learning, basic PBL, and advanced PBL according to a framework of enPiT. Fundamental knowledge learning aims to provide fundamental knowledge relevant to Big data, AI, and Cloud computing technologies to students held in each student's university. Basic PBL is an intensive course style PBL held on throughout 5 days during the summer season. Advanced PBL treats expansive subjects based on basic PBL under the distributed environment. However, because of the difference of each university, it is difficult to align basic knowledge among students through the fundamental knowledge learnings, therefore, we have offered auxiliary fundamental knowledge learning once a month to align preliminary knowledge at least required by PBL. Also, auxiliary fundamental knowledge learning includes corporate seminar too. In the seminar, leading-edge companies

978-1-5386-5605-1/18/$31.00 ©2018 IEEE
DOI 10.1109/BCD2018.2018.00013

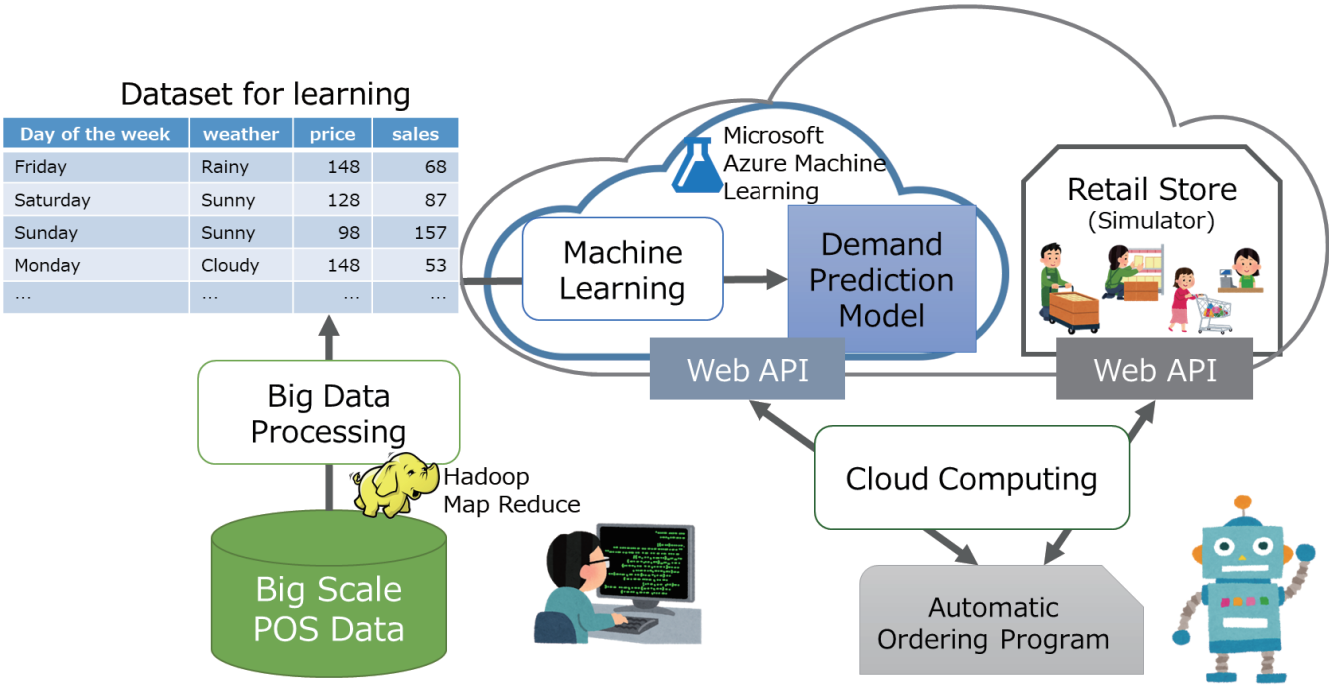

Fig. 1. Overview of PBL Subject

in each fields introduce examples of using in practical fields and its technical account. The goal of AiBiC Kansai is to develop advanced engineers who have knowledge from basic to advanced and practical techniques comprehensively through the lectures, PBL, and seminars.

III. Automatic Ordering System Challenge for PBL

The basic and advanced PBL courses aim to train students so that they can develop advance systems leveraging the latest big data and AI technologies. For the purpose, we have designed an exercise where the students have to develop a system to automate ordering in retail stores.

In the ordering an item, store staffs need to determine an appropriate number of orders so that they can keep the following two losses to a minimum based on the current stock and the prediction of the demand for the item.

- Opportunity loss: the loss of sales opportunities caused by out-of-stock products
- Actual loss: the loss due to discarding or discounting of unsold items

In most retail stores, store staffs order items based on the experience and intuition of the person in charge. However, in our PBL, the students will design a demand prediction model by machine learning using the past sales records obtained from POS data, past weather datasets and so on, and develop a system to automate ordering so that it can maximize the profits of the store (Figure 1).

A. Dataset

The following datasets are prepared in the PBL:

- POS dataset
- Calendar dataset
- Weather dataset

POS dataset includes daily sales records for 116 supermarkets. It includes the number of the sold item and its sales for each day for each store for each item from 2009 to 2013. The total record size is 3.1 billion. Calendar dataset includes information on the day of the week and whether the date is a holiday or not for each day; weather dataset includes weather, temperature, humidity, precipitation and so on for each day.

B. Automatic Ordering System

An automatic ordering system is developed by each student team. The development of the system can be divided into the following two major tasks: development of a demand prediction model and implementation of an automatic ordering program.

1) Development of Demand Prediction Model: The students develop a demand prediction model based on the given datasets using machine learning techniques. The objective variable is the number of sales of an item. The explanatory variables are chosen or developed from the given datasets. Azure Machine Learning Studio[4], [5] (Azure ML) is used to develop the demand prediction model. The created demand prediction model is deployed as a Web service, and can be used by the automatic ordering program via the interface of the Web service. Figure 2 shows an example of the created model on Azure ML.

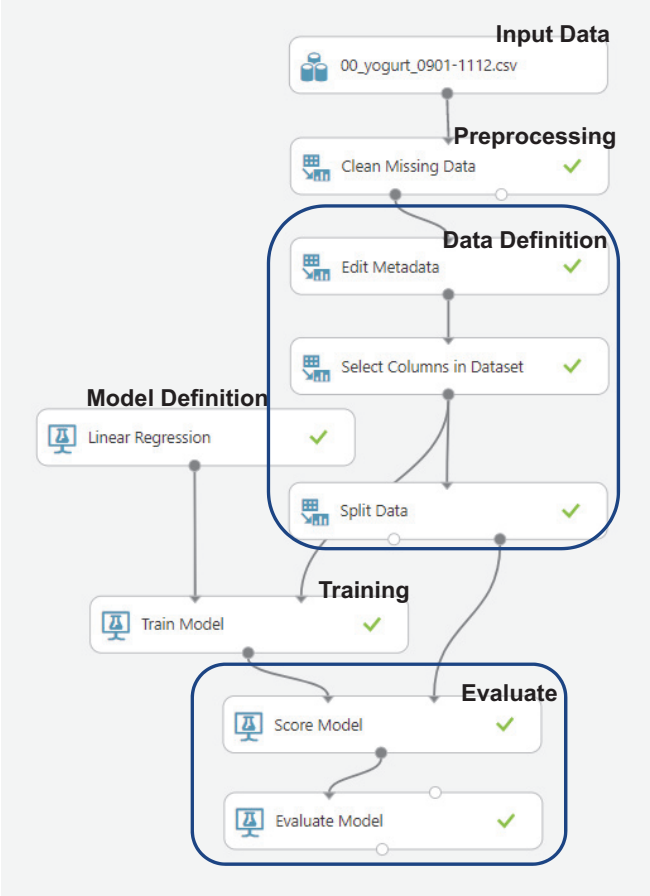

Fig. 2. An example of demand prediction model on Azure ML

2) Implementation of Automatic Ordering Program: Automatic ordering program predicts the demand for tomorrow's sales using the above prediction model and automatically places the orders into a store simulator described later. It determines the number of orders based on the demand predicted by the developed model and the current number of the stocks. The number of orders is not always necessary to the number of differences between the predicted demand and the current stocks. We expect the students to implement some heuristics that takes the characteristics of the target item and the expiration dates of the stocked items into account.

The automatic ordering program is implemented in Python. We provide a template code that places orders into the store simulator. The students add codes for the following functions into the template:

- Calling the demand prediction model via Web service, which is partly generated by Azure ML,
- Generating input datasets to be given to the prediction model,
- Implementing heuristic codes that determine the actual number of orders.

C. Automatic Ordering Competition

One of the goals of the PBL is improving the performance of the automatic ordering system through the competition among the student teams. The performance of the automatic ordering system is calculated by the profit of the store in a specified period. It is hence calculated by the difference between the total sales amount and the total cost of the item during the specified period.

However, just focusing only on the achieved profit does not meet the purpose of the PBL. In particular, if a team achieved a good result by chance, the result cannot be reproduced and the result is not helpful for their future activities.

In this PBL, the contributions of the students can be also evaluated in terms of the following points:

- Selection of the machine learning algorithm
- Tuning of the parameter for the machine learning algorithm
- Selection of the explanatory values for the machine learning
- Implementation of the heuristics to tune the demand prediction

In the PBL, the students are improving the performance of the system by trying the various combinations in terms of the above points. Recording the results of each trial, comparing and analyzing the results of different trials is also important. We therefore define these comparisons and analyses as another goal of the PBL, and ask the students to include these considerations in their final presentations for their evaluations.

1) Store Simulator: If we could evaluate the developed system through a real store operation, it would be very helpful to evaluate the actual performance of the system. However, due to limitation of time and cost of the evaluations, we evaluate the systems with a store simulator that reproduces the behavior of the retail store based on the datasets described in Section III-A. The functions of the store simulator are listed below. The automatic ordering systems access these functions via the REST based APIs.

- Instantiating a store simulator with a specified period
- Placing orders with a specified number
- Retrieving information on the store of the specified date, including the number of sales, stocks and so on
- Retrieving information on the weather of the specified date
- Retrieving information on the weather of the next day
- Retrieving information on the calendar data of the specified date

To simplify the implementations, the store simulator is designed based on the following assumptions:

- The store closes every day (not open 24 hours).
- The store places orders for the next day after closing the store.
- The ordered items will be delivered before the opening of the next day.

TABLE I
ANNUAL SCHEDULE OF 2017

Date	Class
May. 27	Cloud computing
Jun. 10	Big data analysis
Jul. 1	AI
Aug. 5	Integrated study
Sep. 4 - 8	Basic PBL
Oct. 14	Advanced PBL
Nov. 11	Advanced PBL
Dec. 9	Final presentation

- Since we don't have information on the actual cost of each item, the cost of each item is fixed with a specific price in advance.
- Each item has a specific expiration date, and the items passed the expiration date are discarded.
- No discounting based on the expiration dates of the items (The price of the item is decided based on the actual datasets).

When a store simulator instance is instantiated, the date in the simulator is set on the day before the specified period and the store waits for the orders for the next day (it means the orders for the first day). The automatic ordering program can place orders to the store simulator, and then the date in the simulator is forwarded. The ordered items are delivered in the beginning of the next day, and sales and stocks data are updated.

The automatic ordering system can also retrieve the information on stocks, sales, and weather data through the REST API of the simulator. The information is available only for the past data and the current data. As for the weather data, tomorrow's weather forecast is also available. Note that the simulator actually returns the actual data of tomorrow's weather for simplification.

IV. IMPLEMENTATION OF CURRICULUM AND PRACTICAL REPORT

A. Overview

We accepted 52 students from 8 universities and 1 college of technology in 2017. Table I shows an annual schedule. Classes are held at Osaka University Nakanoshima Center. Students learn technologies and facilitation skills for team activities and work on PBL in nine teams consist of five to six students.

B. Auxiliary fundamental knowledge learning

Cloud computing: The cloud computing class focuses on the development of cloud computing technologies including its historical background. This class especially explains virtual machine (VM) technologies, an underlying technology of cloud computing, and students can learn how the flexibility and scalability of cloud computing are achieved by using VM technologies. The class also introduces the big data analysis platform Amazon Elastic MapReduce and the machine learning platform Azure ML. As an exercise, the students construct a virtual machine environment using AWS EC2[6] for further understanding of the advantage of controlling VM resources on cloud computing.

Corporate seminar is held by NTT DATA Corporation and Rakuten, Inc. They introduced the software design tool on the cloud computing and the advantages of using cloud computing for developing a new web service.

Big data analysis: The big data analysis class introduces the definition and applications of big data and explains the MapReduce framework for processing big data. In order to understand data analysis flow, most part of this class are exercises. Before getting into the coding, the students play a mark counting game as a group work. In this game, each group members considered as a worker node of MapReduce. The students then move to coding using Apache Hadoop on the local machine and Elastic MapReduce on AWS cloud computing platform.

Corporate seminar is held by FUJIFILM ICT Solutions Co., Ltd. They introduced the construction of big data analysis infrastructure in the company and application example to the business.

AI: The AI class aims at learning all-around knowledge of artificial intelligence technologies. This class focuses on not only the latest trend such as machine learning but also the history of AI and traditional technologies like an expert system. Since we set the main purpose of this class to understanding the concept and practical use, we particularly consider excluding the detailed algorithm, its calculation formula, and its derivation method. As an exercise, the students work on prediction and classification task on the real estate open dataset. The students implement the task with python and run it on the programming environment on the cloud jupyter[7].

Corporate seminar is held by IBM Japan, Ltd. and The Japan Research Institute, Limited. They introduced IBM Watson and its application in business.

Integrated study : The aim of this class is to learn how to apply cloud, big data, and AI technologies to our automatic ordering PBL. Specifically, the students construct a prediction model on Azure ML with the big data analysis result of POS data. We explain about the evaluation of the prediction model, then explain how to construct a prediction model and evaluate it on Azure ML. After that, the students try to construct a prediction model with various algorithms and parameters.

C. PBL

1) Basic PBL: In the basic PBL course, we first introduce the API specification to handle the store simulator, and each student individually works on implementing a sample program handling the store simulator and the Web service of AzureML. And then, the students work on implementing an automatic ordering system together with the other members in their team. As a practice, the students implement the system for the dataset of the yogurt sales by discussing the points mentioned in Section III-B2 including the selection of machine learning algorithms, tuning of the machine learning parameters and the design of heuristics. For the training datasets, we provide only the first three years datasets of the five years datasets.

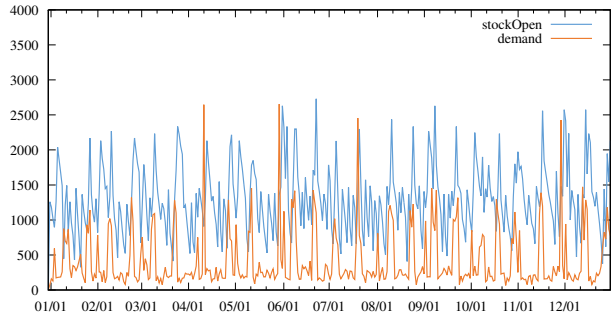

Fig. 3. Basic PBL

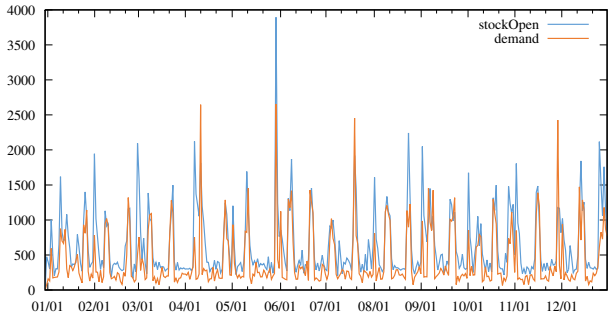

Fig. 4. Advanced PBL

The primary goal of the basic PBL is to let the students learn how to develop a system together with other students in the team. After implementing the first system, the students report their results and the points they have considered as an intermediate report. We also evaluate the performance of the developed systems with the datasets of the fourth year.

2) Advanced PBL: In the advanced PBL course, we provide the datasets for six different items that have different sales characteristics. The students select three items out of the six items, and develop the automatic ordering systems for the selected three items. As mentioned previously, we also make the students record the process of each trial to avoid improving the results by chance without any reasons.

As an indicator to evaluate the prediction accuracy of the system, we use achieved sales ratio which is a ratio of the achieved sales of the developed system to the actual sales recorded in the datasets. Since the sales volumes are different for each item, we use the achieved sales ratio to compare the performance of the systems targeting different items.

At the final presentation, the students in each team give a presentation and report the development activities through the PBL course and the strategies to predict the sales of each item. We also evaluate the performance of the systems using the fifth year dataset in terms of various points including the achieved sales, opportunity loss and the amount of discarded items, and announce the overall ranking.

D. Results

1) Results of PBL: Figure 3 and Figure 4 show a result of automatic ordering for yogurt from the basic PBL and the advanced PBL of a student team respectively. The orange lines indicate the actual demands recorded in the original datasets; the blue lines indicate the number of the stocked item of the retailer. It means that the more similar the two lines are, the more accurately the developed system orders based on the demand prediction.

In the basic PBL, the student team applied decision forest regression as the machine learning algorithm, and implemented the following heuristic process. And, the system has produced the achieved sales ratio of 74.5%.

- If the next day is a closing day, the system does not place any order
- If the sales of today is big enough, the system places more orders for the next day

On the other hand, in the advanced PBL, the student team improved their system and have produced the best achieved sales ratio in the advanced PBL, which is 93.8%. As shown in Figure 4, since yogurt is purchased constantly in everyone's daily lives, it is important to predict the cycle of the sales and order the item efficiently. The result indicates that the developed system places orders quite efficiently. In addition, the number of discarded items of the student team was the smallest in comparison with the other student teams.

This student team applied boosted decision tree regression as the machine learning algorithm, and implemented the following heuristics process.

H1: If the next day is a closing day, the system does not place any order
H2: If the stock of the item is over 200, it reduces the number of the order by 100
H3: If the day with no opportunity loss lasts more than five days, it reduces the number of the order
H4: If today is a closing day and the next is opening day, it increases the number of the order
H5: If the item price of the next day is cheaper than that of today, it increases the number of the order
H6: If the item price of the next day is less than 150 JPY, it increases the number of the order 1.2 times

H1 to H3 are considered as the strategies for reducing actual losses; H4 to H6 are considered as the strategies for reducing opportunity losses.

The improvement of the results from the basic PBL to the advanced PBL indicates that the students have learned about algorithms of machine learning, implementation of heuristics and tuning parameters so that they can develop more efficient automatic ordering system. As the result, the achieved sales ratio has been improved by 19.3%.

2) Evaluation of curriculum by questionnaire: To evaluate our courses in 2017, we conducted questionnaire after the final presentation. We could get 47 answers out of 52 students. Our questions and aggregate results are followings. Fig.5 shows results of questionnaire.

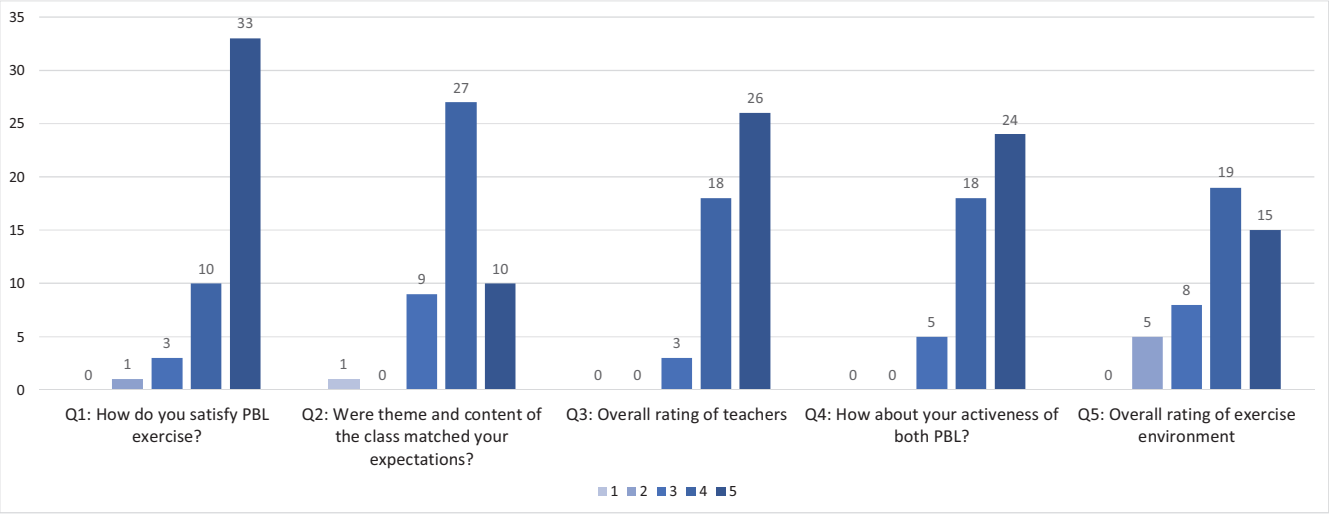

Fig. 5. The results of the questionnaire to the course.

According to the results of the questionnaire survey on courses, from Q1, one student who indicated lack of basic lecture answered negatively, but 43 students were satisfied our PBL. Our PBL was useful for most of the students.

In Q2, while 37 responded that they were consistent, nine students could not say either, one student said that they did not agree. Students who rate less than 3 asked for more lectures on basic knowledge, however In Q1, since more students felt that PBL exercises were beneficial, students who expected to pursue technology deeply also seemed satisfied with this lecture.

From Q3, since 44 people have rated 4 or more, the evaluation for the lecturer is considered high. Therefore, in the PBL exercise management, the lecturer's skill was considered to be sufficient.

In Q4, there were 42 students who responded positively (4 or higher) participation. PBL presupposes student's subjective activities, so It was an indicator of the usefulness of this exercise that getting responses from 89% participating actively participated.

Q5 was a questionnaire for the environment such as class room, facilities, and so on. 5 students responded negatively. They mentioned dissatisfaction to the network environment. Since the exercise was web browser based and network connection was essential, delay of wireless network have occurred by simultaneous accesses of many students. To solve this problem, we have replaced some devices for network connection.

It is difficult to evaluate a curriculum through just one year's result because of the first year of AiBiC Kansai, however, we can get esteem or positive answers from the results of the questionnaire survey in 2017. Furthermore, as comments on the company seminar we get many favorable answers such as "It was a valuable experience that I could listen to the opinions of both the service vender side and the service user side at the same time", "Not only technical aspects but also how they worked normally, what kind of people are working, etc. commentally explained, so I was easy to image the company's work." From these results, our curriculum is producing results to realize the "development of innovative engineer who can address a social challenge using big data processing, AI and cloud computing technologies with the generation of new business and value" as the goal of enPiT AiBiC.

V. Conclusion

In this paper, we introduced our education curriculum named AiBiC Spiral which aims to develop system engineers who can exploit big data, AI and cloud computing technologies practically. The curriculum consists of automatic ordering problem for retail store based on the big data analysis using the machine learning technology. From questionnaire results, we confirmed the educational effectiveness of our PBL centered curriculum. Most of students are satisfied with PBL exercise and its environment. For our future works, we have to improve our curriculum more suitable for educational PBL contents.

Acknowledgment

This education program has been held as a part of enPiT project supported by MEXT. We would like to express special thanks to all members of AiBiC Kansai and students of curriculum (AiBiC Spiral).

References

[1] J. Lee, "A view of cloud computing," *International Journal of Networked and Distributed Computing*, vol. 1, no. 1, p. 2, 2013.
[2] S. Lee, J.-Y. Jo, Y. Kim, and E. Hwang, "Big data analysis with hadoop on personalized incentive model with statistical hotel customer data," *Int. J. Softw. Innov.*, vol. 4, no. 3, pp. 1–21, Jul. 2016.
[3] enPiT AiBiC, https://aibic.enpit.jp/.
[4] Microsoft Azure Machine Learning, https://azure.microsoft.com/ja-jp/services/machine-learning/.
[5] R. Barga, V. Fontama, W. H. Tok, and L. Cabrera-Cordon, *Predictive analytics with Microsoft Azure machine learning*. Springer, 2015.
[6] J. Varia, "Architecting for the cloud: Best practices," *Amazon Web Services*, vol. 1, pp. 1–21, 2010.
[7] Project Jupyter, https://jupyter.org/.