DS&AI

Co-funded by the Erasmus+ Programme of the European Union

Curriculum Development in Data Science and Artificial Intelligence

599600-EPP-1-2018-1-TH-EPPKA2-CBHE-JP

Identification of curricula in the area of Data Science and Artificial Intelligence in Asia

We teach today's students
With yesterday's knowledge
For a future we don't know.

Version 1: 15 August 2019

DS&AI

Co-funded by the Erasmus+ Programme of the European Union

PROJECT INFORMATION

Acronym	DS&AI
Project Title	Curriculum Development in Data Science and Artificial Intelligence
Contract Number	599600
Start Date	15 November 2018
Duration	36 months

DELIVERABLE INFORMATION

Deliverable Number	D1.1
Deliverable Title	Report on similar curricula in Asia
Submission Due Date	14 March 2019 (proposal) 15 August 2019 (revised)
Actual Submission Date	15 August 2019
WP Number and Title	WP1: Preparation
Lead Beneficiary	WU (originally PSU)
Dissemination Type	Report
Dissemination Level	International
Quality Reviewer 1	AMC
Quality Reviewer 2	WU
First Quality Review Date	12 February 2020
Quality Review Pass Date	31 May 2020

DISCLAIMER

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Abstract

In this report we survey and identify existing master programmes in Asia in data science and artificial intelligence. Additionally, I reports on focus group discussions organized in each of the partner countries (3 in total) on needs and gaps that have to be addressed by this project. The focus group were composed of several academics, students and representatives from ICT companies in order to guarantee that all relevant target groups provided their input in developing a master curriculum on data science and artificial intelligence.

CONTENTS

1. Introduction

The average introduction to almost any book is somewhat of a bore

--- Boris Karloff

Across many industry sectors, including retail, agriculture, healthcare, banking, finance, and telecommunications, companies in South Asia are launching initiatives to improve their operations and offerings through data-driven techniques[1]. One of the main success factors in this digital transformation is having the right talent with critical skills in data science and artificial intelligence, with the ability to understand both data processing techniques and the business they are operating in.

1.1 Context

Thailand, Indonesia, and Sri Lanka are middle-income countries within the South Asia region. While data science and artificial intelligence based technologies are gaining traction and becoming strategic priorities these countries are still falling behind regional leaders in the technological market, such as Singapore and Malaysia, while the region as a whole is far behind China. Thailand, Indonesia, and Sri Lanka have well-established and growing higher education systems; increasingly focused on quality improvement. However, they fail in educating people who design, develop, deploy, and train data science and artificial intelligence-based technology, as evident from the acute shortage of data analysts in public and private companies[2,3].

These trends make clear that development of talent and capabilities is needed if data science and artificial intelligence are to reach their full potential throughout the region. The success of political and economic initiatives by governments in the south Asian region to transform the current export based economy into an innovation and knowledge based economy will critically depend on the readiness of the labour force for ICT in general and data-driven technologies in particular.

This document focuses on the identification of needs and gaps in master level education in data science and artificial intelligence through South Asia. It aims to provide an account of a common core of what a Data Science and Artificial Intelligence master programme should provide, and how this core can be extended to satisfy the specific needs of countries of Asian partners.

---

[1] *The age of analytics: competing in a data driven world*, McKinsey Global Institute, 2016.

[2] *Artificial intelligence and South-East Asia's Future,* McKinsey Global Institute, 2017.

[3] Mathur, Aneja, Anwar, Shridhar, and Sanchez, *Future of Work in Sri Lanka: Shaping technology transitions for a brighter future*, TANDEM research report, International Labour Organization, pp. 1—98, 2019.

Data science is concerned with the extraction of useful domain knowledge from large, complex data sets. Drawing on the fields of machine learning and data mining, it provides critical support to decision makers in many professions, allowing them to make decisions based on statistically significant patterns in data. Applications include business intelligence, the detection of anomalies in patterns of behaviour, and for example the analysis of services in cloud computing.

Artificial intelligence is an even broader field concerning the theory and development of computational systems able to perform tasks inspired by human intelligence. We can distinguish three main paradigms within the field of Artificial Intelligence:

1. *symbolic AI* – consisting, among others, of inductive logic programming, robotic process automation and expert and fuzzy systems,
2. *statistical AI* – consisting of decision networks, probabilistic programming (including Bayesian program synthesis), computer vision (activities and image recognition as well as machine vision), natural language processing, and machine learning, and
3. *subsymbolic AI* – consisting of distributed artificial intelligence (including agent-based modelling, swarm intelligence and multi-agent systems), ambient computing, natural computing, multi-objectives optimization, affective computing, embodied intelligence, and autonomous systems.

Today the most widely applied discipline within AI is machine learning, a method of data analysis that automates analytical model building without being explicitly programmed. Machine learning comes in at least three "flavours": supervised learning, unsupervised learning, and reinforcement learning. Supervised learning refers to methods for learning a model from a set of training data containing both input and output results. Regression and classification are typical examples of supervised learning techniques. Unsupervised learning consists of building a model from a set data, via for example visualization or clustering. Reinforcement learning is about taking actions in an environment so as to maximize some notion of cumulative reward, as for example used in robot navigation and games (including chess, Go, and business games).

1.2 Methodology

The high-level methodology of the project contains of five steps, each described in a separate chapter of the report. The first step is about **identifying academic curricula on data science and artificial intelligence in the Asian partners' countries**. Through desktop research and using a common template we have collected data from several master programmes on Data Science and Artificial Intelligence in the regions of Thailand, Sri Lanka, India, Malaysia and Indonesia. Many of the programmes we collected data from are generic masters in computer science and contain only few courses related to data science or artificial intelligence. Only 13 out of 35 programmes we investigated are directly related to data science or artificial intelligence. Of these only two are full-fledged master programmes on artificial intelligence. We complemented the information about the programme with a discussion of the credit system in Asia.

The second step consists of **identifying the gaps that have to be addressed in the development of master's degree programmes in data science and artificial intelligence** in Thailand, Indonesia and Sri Lanka. With this goal, partner universities in each of these countries organized a focus group composed of 4 academics, 3 students and 3 representatives from ICT companies. This composition of the focus group guaranteed that all relevant stakeholders provided their input in understanding what is needed by a new education programme in data science and artificial intelligence.

2. DS and AI Master Curricula through South Asia

Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom

--- Clifford Stoll

2.1 Introduction

The use of systems based on data science and artificial intelligence techniques is rapidly growing in several sectors around the world, e.g., in the health care, transport and financial sectors. This revolution is accelerated by advances in data collection, storage and analysis, efficient algorithms, and an increasing computer processing power. Although the United States, China and Europe are the frontrunners in developing such artificial intelligence systems, these world-wide scientific advances also have a major impact on countries from Southeast Asia. As the fields of data science and artificial intelligence have the potential to contribute positively to the economic and social climate of those countries, there seems to be a great need for training professionals in this area.

The aim of this chapter is to identify existing master level curricula related to Data Science and Artificial Intelligence in Thailand, Indonesia, Sri Lanka and wider region. While in Europe modern higher education programmes have been developed on specific areas of computer science such as data science, data analytics, business analytics and artificial intelligence system, in many countries from Southeast Asia computer science education programmes are often developed in a generic way, e.g., based on ACM curricula recommendations. Most common master programmes are therefore named Computer Science, Information Technology, Information System, and Computer Engineering. These programmes often include several elective tracks, such as hardware and computer architecture, communication and networking, computer system or software. But there are no programs specific to data science and artificial intelligence. Often, students are only allowed to follow only a limited number of elective courses within their track, as such the postgraduate knowledge of these students cannot be compare with the postgraduate knowledge of European students who followed specific master programmes dedicated on data science or artificial intelligence.

2.2 Education and credit system in Thailand, Indonesia and Sri Lanka

In Asia, universities often offer two kinds of masters' degree: coursework-based and research-based. Research-based master programme does not require any course work and the admission is usually based on a specific bachelor's degree. Coursework-based master programmes requires to gather information by taking courses. In many universities, students must perform an additional independent study project or an independent research project, after their coursework. Coursework-based programmes are typically

shorter than the research-based programmes. The coursework-based programmes take 1-2 years while research-based programmes takes between 2-3 years. In Asia the nominal duration of a masters' degree is typically two years.

An academic year is usually divided into the first and second semester, they may rarely be an optional third summer session. Each semester length is 16-18 weeks and include an examination. Prospective student can apply for masters' degree twice a year. A number of minimum credits are between 36-45 in South-East Asia and nearly 60 in South Asia.

Each university in Asia is managed by its own polices and regulations, with the only constraint to be consistent with governmental strategies and policies. Quality assurance is usually supervised by government organization such as University Grant Assurance (UGC), Quality Assurance and Accreditation Council (QAAC), High Education Institute or Minister of Education. Some programmes have professional council to guarantee quality as required by the labour world (e.g., for the fields of engineering, medical or pharmacy). In Asia there is no Data Science and Artificial Intelligence council to guarantee quality of higher education.

Asia Credit Transfer System (ACTS)

A credit system in higher education is widely accepted as a way of describing educational programmes by attaching credits to its components. The credit value's role is to indicate the average amount of learning time required for a learner to achieve a credit value. Learning time is made up of lecture-based learning as well as independent learning, e.g., self-learning online, doing homework and research and work placement. There is not a unique standard for academic credit unit and accreditation system. Universities in the Association of South East Asian Nations, or ASEAN, member states utilize a credit transfer system. Traditionally, recognition of semesters abroad is been carried out on a case-by-case basis. When students study aboard, they must transfer credit into other credit system such as ECTS. Asian credit earning systems generally are defined as follows

- Lecturing or any academic activities that are equivalent to 1 hour per week throughout the whole semester or approximately 15 hours in a semester is measured as 1 credit
- An operation, experiment, or a laboratory activity that takes 3 hours per week throughout the whole semester is measured as 1 credit.

Although currently no global credit transfer system for the ASEAN region exists, certain systems are used to help streamline the process. In order to realize a student-centered higher education system in Asia, the Asian Cooperation Dialogue recently developed an Asia Credit Transfer System (ACTS) based on the successful three decades of European experience with the European Credit Transfer System (ECTS). ECTS uses 60 credits as standard for one year of study or 30 credits for each semester or 20 credits for each trimester. A typical master's degree would consist of 60 (one year) to 120 (two years) credits, and each credit is based on 28 hours of work for the student (based on the average working time of 1560 hours per year in Europe). ECTS is a very systematic and permeable credit transfer system. However, student workload for ECTS does not reflect an Asian workload (1560 hours per year in Europe vs. 1800-2100+ hours per year in Asia). Therefore the ACTS coefficient depends on the minimum number of credits in

each programme, and it is not standard even within the same country. In general, Asian master's programs have a duration of two years and they normally require 36-45 credits in total. Therefore, one credit in a master's degree is approximately 2.6 to 3.3 ECTS. For example

- 1-year study in Europe = 60 ECTS (standard)
- 1-year study at PSU in Thailand = 36 credits (the minimum required credits in total) / 2 years (the duration for master's degree) = 18 credits

Therefore, the coefficient to convert 18 credits to 60 credits is 3.33 ECTS per local credit. As such, a course at PSU consisting of three credits is equivalent to 10 ECTS.

2.3 Relevant master programmes in Thailand, Indonesia and Sri Lanka

This section focuses on the learning knowledge related to existing master programmes in data science and artificial intelligence in Thailand, Indonesia, Sri Lanka and wider region. The knowledge topics have been grouped by the course type which can be divided into three groups including core, elective and supportive course. The core courses are the mandatory courses in the curricula. The elective course are additional course that can be selected by students among several alternatives. The supportive courses are related to the basic knowledge which may be the pre-requirement for admission or a pre-master course or a provided course in the programme. Supportive course may construct under some condition such as academic English language or basic mathematic and statistic. There are five groups of core courses, i.e.:

- Business Intelligence (data modeling, data mining, data analytics and data visualization);
- Data science (programming, data analytics, data warehousing, management and cloud computing);
- Machine Learning ;
- Mathematics and Statistic (theory and methods, regression analytics, statistical analysis and statistical inference);
- Research area (research methodology, independent study, seminar and project).

The most popular topics among the core course are data management, data mining, data warehousing, machine learning and research methodology. There are various elective topics which are hard to categorize, because the non-uniform naming of courses. There is only one topic which is taught in all programmes: data and text analytics. Finally, all programmes make available three groups of supporting knowledge: programming, algorithmics, and mathematics and statistics.

2.3.1 Thailand

The fields of data science and artificial intelligence is currently a part of national strategic plan of Thai government, called Thailand 4.0. Thailand 4.0 is a 20-year strategy to accelerate the country's development from upper-middle-income country to high-income country status. The strategy is designed to promote and support innovations, creativities, research, higher technologies, and green technologies. Thailand 4.0 is correlated to global Industry 4.0 and focuses on 10 industrial areas, i.e., next-generation automotive, intelligent electronics, advanced agriculture and biotechnology, food

processing, tourism, digital, robotics, logistics, biofuels and biochemicals, and medical. The goal is to make possible a digital transformation of manufacturing by combining a wave of next-generation technologies, and thus heavily rely on the advancement and technologies from the field of data science and artificial intelligence. Therefore, the development study programs in these particular fields and the related areas are encouraged by the government.

Beside the formal approval of a new higher education programme on data science and artificial intelligence, the major challenge comes from the northern and north-eastern regions of Thailand, because of the limited number of experts who are eligible to teach and supervise research in master and doctoral programs. The solution is a dedicated development plan and training helping in recruiting more expert teaching staff

Next we summarize the curricula of existing master programmes in areas related to data science and artificial intelligence. These programme are offered by public universities in the northern and north-eastern regions of Thailand. All information was collected and analysed through desk research and interviews.

Master of Science Program in Business Analytics and Data Science

*By:*National Institute of Development Administration, Graduate School of Applied Statistics.

Duration: 2 years with 45 minimum credits.

Core topics:

- Mathematics of Computing (Discrete Math, Probability and Stats)
- Data Modeling and Management
- Data Mining and Information Retrieval
- Natural Language Processing
- Graph Theory and Network Science (social network analysis, complex networks)
- Applied Data Science and AI
- Human-Computer Interaction and Information Visualization

Elective topics:

- Language and Speech Technology (Voice Recognition)
- Machine Learning
- Data and Text Analytics
- Software Engineering (for DS&AI Systems)
- Applied Data Modeling and Management

Support knowledge: Background knowledge in programming, algorithms, mathematics, statistics and data communication (computer network) is required. For those without related background, preparatory courses are provided.

Master of Science Program in Data and Information Science

By: Rajamangala University of Technology Thanyaburi, Department of Mathematics & Computer Science.

Duration: 2 years with a minimum of 36 credits.

Core topics:

- Mathematics of Computing (Discrete Math, Probability and Stats)
- Data Modeling and Management

- Human-Computer Interaction (Visualization)

Elective topics:

- Context of Artificial Intelligence (History, Philosophy, Ethics)
- Data and Text Analytics
- Machine Learning
- Data Mining and Information Retrieval
- Internship

Support knowledge: Previous degree in computer engineering, software engineering, computer science, information technology, business computing, mathematics, statistics or related disciplines, or related working experience of at least 5 years is required.

Master of Science Program in Data Science

By: Srinakharinwirot University, Department of Computer Science.

Duration: 2 years with a minimum of 37 credits.

Core topics:

- Data Modeling and Management
- Machine Learning
- Cloud Computing

Elective topics:

- Applied Data Science and Artificial Intelligence
- Data and Text Analytics
- Image and Video Analytics
- Bioinformatics
- Security and Privacy
- Graph Theory and Network Science (social network analysis, complex networks)

Support knowledge: Background knowledge in programming, algorithms, mathematics and statistics is required. For those without related background, preparatory courses are provided.

Master of Science (Data Science)

By: Chiang Mai University (CMU).

Duration: 2 years with a minimum of 36 credits (24 coursework and 12 thesis, or 30 coursework and 6 independent study).

Core topics:

- Statistics for Data Science
- Data Management and Big Data
- Data Science Programming
- Everything Start with Data

Elective topics in study area: Analytics and e-Commerce

- Business Data Analytics
- Data Analytics for e-Commerce
- Data-driven Customer Relationship Management
- Data Methodology for Web Search Optimization
- Selected Topics in Business and e-Commerce 1 & 2

Elective topics in study area: Data Engineering and Architecture

- Advanced Embedded Systems
- Parallel Processing and Distributed Systems
- Information Security
- Database Management Systems
- Information Technology Infrastructure Library
- Software Project Management
- Data Governance
- Cloud Computing
- Selected Topics in Data Engineering and Architecture 1 & 2

Elective topics in study area: Data Analysis and Machine Learning

- Data Analytics and Machine Learning
- Artificial Intelligence
- Operational Research Techniques
- Statistical Decision Method
- Demographic Statistics
- Methods of Statistics
- Computer Packages for Advanced Statistical Analysis
- Forecasting Techniques
- Numerical Analysis
- Selected Topics in Data Analysis and Machine Learning
- Statistical Quality Engineering and Control

Support knowledge: Background knowledge in Science, Engineering, or other related fields with good Mathematics background or have work experience in the data science fields.

2.3.2 Indonesia and wider region

The need for higher education in the field of data science and artificial intelligence has triggered the development of higher education programmes in these areas. However, the government of Indonesia has some regulations related to the naming of a study programme. In Indonesia, before a new higher education study programme can be established, it should be first accredited by the National Accreditation Board of Indonesia. The name of the study programmes in the computing field must be as specified in the ACM Computing Curricula: Computer Science, Information Technology, Information Systems, and Computer Engineering.

Due to the above naming regulations, data science and artificial intelligence programmes are typically presented under existing computer science, information technology, and Information systems curricula. For this reason below we summarize only those curricula of existing master programme the areas of data science and artificial intelligence in the neighbourhood country Malaysia.

Master of Science Program in Data Science
By: Universiti Kebangsaan Malaysia.
Duration: 1 to 2 years with a minimum of 40 credits.
Core topics:

- Data science fundamentals

- Statistical Methods
- Structured data analytics
- Unstructured data analytics
- Information modeling and database
- Big data analytics and management
- Machine learning

Elective topics:

- Business Intelligence
- Executive Information Systems

Support knowledge: Background knowledge in programming, algorithms, database and data warehouse fundamentals.

Master of Data Sscience

By: University of Malaya.

Duration: 2 years with a minimum of 42 credits.

Core topics:

- Research Methodology
- Principles of Data Science
- Data Analytics
- Programming for Data Science
- Data Mining
- Machine Learning for Data Science
- Big Data Management

Elective topics:

- Parallel and Distributed Computing
- Big Data Applications & Analytics
- Network and Security
- Numerical Optimization

Support knowledge: Background knowledge in programming, algorithms, and methods in artificial intelligence.

Master in Data Science and Business Analytics

By: Asia Pacific University of Technology and Innovation (APU) – Malaysia.

Duration: 12 months over 3 semesters of 12 weeks each.

Core topics:

- Big Data Analytics & Technologies
- Behavioural Science, Social Media and Marketing Analysis
- Data Management
- Business Intelligence Systems
- Research Methodology
- Applied Machine Learning
- Data Analytical Programming
- Multivariate Methods for Data Analysis

- Capstone Project 1
- Advanced Business Analytics and Visualisation
- Capstone Project 2

Elective topics:

- Time Series Analysis and Forecasting
- Natural Language Processing
- Operational Research and Optimization
- Multilevel Data Analysis
- Strategies in Emerging Markets

Support knowledge: Background knowledge in statistics, programming, artificial intelligence methods, management.

2.3.3 Sri Lanka and wider region

The situation of higher order education in the fields of data science and artificial intelligence in Sri Lanka is very similar to the one in Thailand, with a government pushing towards a new digital revolution in industry but with a not well developed higher order education programmes specialised in data science and artificial intelligence to absorb the increasing request from the labour market in these areas. Below we summarize the curricula of the current master programmes in data science and artificial intelligence in Sri Lanka and wider region.

Master of Business Analytics

By: University of Colombo, School of Computing.

Duration: 2 years with a minimum of 22 credits in first two semesters.

Core topics:

- Business Statistics
- Organizational Data Management
- Fundamentals of Business Analytics and Data Science
- Data Programming
- Statistical Inference for Analytics
- Machine Learning and Pattern Recognition
- Data Warehousing and Mining
- Information Visualization
- Modelling and Simulation of Data
- Predictive Analytics
- Computational Social Sciences

Elective topics:

- Applied Optimization
- Open Source Intelligence
- Text Analytics
- Project on Business Analytics
- Independent Studies in Business Analytics
- Big Data Analytics
- Analytics for Process Improvement

- Intelligent Agents in Gaming

Support knowledge: English.

Master in Artificial Intelligence

By: University of Moratuwa.

Duration: 2 years with a minimum of 56 credits.

Core topics:

- Programming Essentials for Artificial Intelligence
- Essentials of Artificial Intelligence
- Mathematics for Artificial Intelligence
- Distributed Computing Concepts for AI
- Deductive Reasoning and Logic Programming
- Neuroscience & Neurocomputing
- Evolutionary Computing
- Artificial Cognitive Systems
- Fuzzy Reasoning
- Data Mining and Data Warehousing
- Software Agents and Swarm Intelligence

Elective topics:

- Cryptography and Security Mechanisms
- Semantic Web and Ontological Engineering
- Intelligent Solutions for Industry
- Natural Language Processing
- Embedded Robotics
- Inductive Logic Programming
- Kansei Systems

Master in Computer Science, specialization Data Science, Engineering, and Analytics

By: University of Moratuwa.

Duration: 2 years with a minimum of 60 credits.

Core topics:

- Data Mining
- Big Data Analytics Technologies
- Data Science
- Business Intelligence
- Machine Learning
- Statistical Analysis
- Statistical Inference
- Advanced Algorithms
- Neural Networks
- Research/Industry Projects

Master Business Analytics

By: Informatics Institute of Technology (IIT).

Duration: 2 years.
Core topics:

- Data Management
- Introduction to Big Data and Data Science
- Business Intelligence Tools and Applications
- Statistics for Business Analytics
- Business Modelling and Analytics
- Fundamentals of Data Warehousing
- Research Methods
- Web Mining
- MSc Project

Master Big Data Analytics
By: Informatics Institute of Technology (IIT).
Duration: 2 years.
Core topics

- Advanced Databases
- Big Data Programming
- Cloud Computing
- Data Analysis
- Data Mining
- Data Warehousing
- Research Methods
- Text Analytics
- MSc Project

Master of Technology, specialization in Artificial Intelligence
By: Indian Institute of Science – Bangalore
Core topics

- Digital Signal Processing
- Machine Learning
- Pattern Recognition
- Computer Vision
- Graphics
- Stochastic Systems
- Multimedia
- Real Time and Fault Tolerant Systems
- Sensor Networks
- E Commerce
- Speech Processing
- Real Time Signal Processing and Embedded Systems.

Support knowledge: C and C++

2.4 Conclusions

Currently, data science and artificial intelligence are popular topics around the world. Many universities in Asia have specialised master's program to support the need for the labour market. However, there are several problems in current programmes. Firstly, some universities are regulated by governmental rules about naming, therefore, data science and artificial intelligence specialised programmes must be encapsulated under a general programme of computer science according to the ACM curricula recommendations. Secondly, there is no standard set for data science and artificial intelligence topics, therefore, current curricula have been designed by the present expertise within each university. Thirdly, most of the current master's programmes do not have internships as part of their curricula, creating an unnecessary distance between the labour and the academic world.

3. Gaps Analysis in Master Curricula through South Asia

If you bend-over to analyse a gap too long, you'll probably fall into it.
--- Ryan Lilly

3.1 Introduction

Economic and social development increasingly depend on innovation. Universities have a potentially important role in driving innovation and development. They can do so both through their role in carrying out research and development and by training workers for the knowledge economy. In the previous chapter we have seen that companies in the South Asian region have a great demand of employees in the field of data science and artificial intelligence. The need for a master degree in these fields is vital for companies to grown larger and keep pace with the advancement in the rest of the world, more importantly in USA, China and Europe. Higher education is seen to have an ever more important role not only in human resource development but also in bringing students to companies through internships and in the regional workforce education via professional training.

Higher education in computer science across South Asia faces new challenges: maintaining and improving education quality, even in the face of serious financial constraint; improving the relevance of curriculum and instruction at a time of rapid change in labour market needs (as we explored in the previous chapter). Extraordinary effort needs to be made in diversifying curricula, and experimenting with new instructional delivery systems, as discussed in the European context in Chapter 2.

In view of the companies' needs analysis of Chapter 5, based on the current situation of the master education in data science and artificial intelligence as sketched in Chapter 3, and using the experience of similar European programmes as discussed in Chapter 2, in this chapter we present the result of a gaps analysis. We used a focus group discussion to identify needs and gaps that have to be addressed in the development of Master's degree programme in data science and artificial intelligence in Thailand, Indonesia and Sri Lanka. In particular, the following issues and questions are investigated:

Issue 1: Identification of subject areas related to data science and AI that are most in need in the country (and wider region);
Issue 2: Identification of a set of skills that the curriculum should develop and promote;
Issue 3: Teaching and learning processes that are appropriate for the curriculum (project-based learning, professional certifications, trainings, practical and industrial projects, workshops, internships, research, theses, etc.);
Issue 4: Required resources, facilities, tools, as well as support that the universities should provide (what we lack or need)

Issue 5: Any concerns and opinions regarding internships and job opportunities that the curriculum should provide

The discussion comprises at least 15 participants from three groups, namely academics (7 or more), ICT industrial sector (4 or more), and students (4 or more), and hence allowing us to obtain significant inputs in developing a balanced Master's degree programme. The participants were asked to participate in a discussion and share their inputs and opinions regarding the development of data science and artificial intelligence curricula, including their opinion on job and skill demands, needs for new courses, pedagogy and instructional methods, internships, job opportunities, required tools, resources and facilities. Specifically, the five issues mentioned above were discussed in detail. Information about the focus group participants for each country can be found in Appendix A.

3.2 Focus group discussion Thailand

This section presents the findings from the focus group discussion in Thailand from 3 different perspectives: ICT industrial sector, student and academics, respectively.

Issue 1: Identification of subject areas related to data science and AI that are most in need in Thailand (and wider region)

ICT Industrial Sector Perspective:

- Jobs with currently strong need and most preferable by the industry sectors are Data Engineers. They should have the abilities to perform ETL (Extract Transform Load), Data warehouse manipulation (Big data repository), and data preparation skills which are found lack at present. Currently, many companies employ data scientists working as a role of data engineers. Data Scientists are normally expected to use the latest data technologies so as to solve business problems for strategic advantages.
- The curriculum should provide both fundamental and advance core courses composing of SQL, NoSQL, Data Ingestion, DDL, Data Warehouse Logic, Programming, SQL for Hadoop big data platform. ICT Industrial deal with bank sector proposed Probability and Statistics, Programming in Python, Linear Algebra. Machine Learning, Neural Network, Deep Learning, Genetic Algorithm, Text Mining and IR, Data Structures and Algorithms, NLP (Natural Language Processing) and Text mining are also essential to learn.
- One thing that should be noted is that students should understand how to utilize when applying Machine Learning algorithm in their works by using ready-made libraries or off-the-shelf component. These useful tools can lead to further development as a product such as Python, Hadoop, Matlab, and Weka are needed to learn in order to improve capabilities of the products. Students should enhance Business Analysis or Market Analysis which will certainly be used in real work. In practice, working together between the Business user and

Data scientist will have a greater chance of success rather than trying to get Data scientist to understand Business on their own.

- Start-up companies usually require all tasks in one employee with wide and various knowledge, such as business, legal, marketing, and the ability to produce products. Machine Learning, Deep Learning and Genetic Algorithm
- Data governance also should be taught in this program due to concerning the capability that enables an organization to ensure that high data quality exists

Student Perspective:

- Basic subjects such as Data modelling and management (Data structure, Data Pre-processing) and more advance courses related to Big Data, Data mining, Mechanical learning and Robotics are expected to be taught according to student needs. The special topics should be brought from industry-based projects on different domains to help student better understand the core concepts of Data Science and Big Data.
- Students should be given the opportunity to choose the subjects that interest them. Moreover, teaching courses should not specificied in some particular domestic job markets so that students can find jobs worldwide with higher income.

Academics Perspective:

- Teaching and learning should lay down the foundation of core knowledge together with the ability of using the various tools and software i.e. Python that will be implemented to develop the product in practical. Moreover, collaboration with industrial section can lead to students' better learning by applying different real case studies.
- The example of important fundamental courses recommended by lecturers are programming development such as Algorithm, Logic, Data structure, Programming and Data management such as Database / SQL. - SQL, NoSQL. In addition, Discrete mathematics, Linear algebra, Probability, Statistics, Operation research are referred by ICT industrial that should be included into core courses such as Introduction to Data Science, Machine Learning, Data Mining and so on.
- The target groups of learner may be various depending on learner's objectives and business companies involved in that region i.e., some learners with their own data/information will have a desire to learn how to analyse; they focus on business analysis and require the algorithm at the basic level. Others may focus on in-depth algorithms for the maximum efficiency product customization and development and these in-depth learning should be also leaded to research capabilities as well.

Issue 2: Identification of a set of skills that the curriculum should develop and promote

ICT Industrial Sector Perspective:

- The priorities of skills recommended from the industrial sector are communication, analytical skills, empirical skills, presentation and Argumentation, respectively. Moreover, Data visualization (Infographics) should be taught to effectively present all analyzed results to customers or relevant users. In addition, Teamwork, Problem solving skills and Attitude are important as well.

Student Perspective:

- According to students' points of view, the communication skills are very important. systematic thinking, effective communication and problem solving should be taught through this curriculum.

Academics Perspective:

- The important skills that student should gain through this curriculum are an in-depth understanding, Analytical skills, Empirical methods, Modelling, Teamwork, Written and oral communication, and Argumentation and presentation.

Issue 3: Teaching and learning processes that are appropriate for the curriculum (project-based learning, professional certifications, trainings, practical and industrial projects, workshops, internships, research, theses, etc.)

ICT Industrial Sector Perspective:

- According to industrial sectors' points of view, students should gain the practical skills through working on real world problems in an industry collaborative environment through projects. Project base learning should be taught due to it is a dynamic approach solving real-world problems to gain knowledge and skills. Through this learning experience, students will be able to investigate and respond to an engaging and complex question, problem, or challenge.

Student Perspective:

- Students prefer continuous project assignments to typical study. In addition, assignments should suit each student's ability. Workshops and invited speakers should be organized in order to provide students usable skills, and recognized qualifications.
- Industrial internship is also vital in order to boost student's practical skills while doing the projects with industrial partners.
- A variety of tools applied in industry nowadays should be provided during the teaching and learning process.

Academics Perspective:

- The teaching and learning process should offer two different tracks: Industrial project/Internship track and Research track.
- Students should experience to practice by using practical problems from industries. In case of big data study, there should be enough dataset for doing experiments. Therefore, it is very

important for the university to have collaborations with various industries leading to better teaching and learning.

- However, the business confidentiality is a sensitive issue. This leads to the cooperative research agreement between the university and industrial sectors instead of requesting data set /information used for teaching and learning.

Issue 4: Required resources, facilities, tools, as well as support that the universities should provide (what we lack or need)

ICT Industrial Sector Perspective:

- Instead of paying for expensive software in learning, students may suggest to use some free software or education version from various vendors such as AWS educate, Google cloud for education, Hadoop, Python, Spark cluster, Cloudera, Tensorflow, etc.
- Seeking strength in partnerships and expanding research capabilities, the collaboration ecosystems should be well planned.
- The start-up labs and product show case should be set to meet the entrepreneurial requirements and must be equipped with laptops, database, ecosystem (AI big data), licensed software and necessary things.

Student Perspective:

- Inviting speakers from industrial sectors to give a talk can provide students more relevant ideas.
- The laboratory equipment should be prepared and ready to use for the most effective learning such as hardware resources for high computation, Machine learning, Neural Network and Parallel Processing.
- The tools used in teaching and learning should be a tool that is in line with those commonly used in the industry.

Academics Perspective:

- To practice data processing in big data experiments, the high-performance laboratory equipment should be prepared and the tools should be suitable for learning various algorithms both basic and advance level. Academics proposed that ICT Industrial should cooperated and support with university to provide up-to-date tools.

Issue 5: Any concerns and opinions regarding internships and job opportunities that the curriculum should provide

ICT Industrial Sector Perspective:

- Inviting experts from companies is also a good idea to inspire students with valuable practical knowledge focused on graduate ready skills for future careers.

- Students should be encouraged to join various community groups in order to update new knowledge and modern tools. Moreover, students should have a chance on some practical works related to the special topic that they are interested in.
- In order to apply for a job, join an internship among other competitors, a student portfolio is also important because that reflects all accomplishments, skills, experiences, and attributes. Students should learn how to professional create.

Student Perspective:

- Students expect to have the Industrial Training/Internship opportunity practiced by private companies/organizations so that they can understand organizational environment and perform well towards organization objectives.

Academics Perspective:

- Industrial Internship is good way to give opportunity with real case study for student's practice, which job opportunities in data science and artificial intelligent are needed to solve to explore insight data and can be applied how to get business value following the ICT's requirement. However, the academics concern about the taking course time that should be combined to research work.

Conclusions

The teaching and learning of the data science and artificial intelligence curriculum must be able to meet both the needs of the industry and the interest of students. Teaching and learning should lay down the foundation of core knowledge that is sufficient for learning data science and artificial intelligence, including (1) Mathematics and statistics groups such as Discrete mathematics, Linear algebra, Probability and Statistics, Operation research (2) Programming development groups such as Algorithm, Logic, Data structure, Programming and (3) Data management groups such as Database and SQL.

The need for knowledge of Business analytics and Machine Learning depends on whether basic or in-depth learning depending on the student group characteristics and the objectives of the student's study. Therefore, the developed curriculum should have different tracks to be chosen between (1) Applying data analytics techniques to solve problems, analysing data for the industry and (2) Researching to develop deep knowledge in DS / AI science, which leads to enhance the higher performance for the particular industry.

The laboratory equipment should be prepared and ready to use for the most effective learning. The tools used in teaching and learning should be a tool that is in line with those commonly used in the industry. Students should have the skills to use those tools that can lead to the product development, not just a tool that running or testing as experiments. The course must be built up the collaboration with the industry in applying the real case study for teaching or for students to experiment. In particular, students should practice as internship in a company, or do some industrial projects.

Not only hard skills that students can gain through education, training programs and internship training, but also soft skills for data science and artificial intelligence students are very important. The soft skills

that are suggested to be developed including (1) Communication skills, (2) analytical skills/ empirical skill, and (3) Presentation skills, both in terms of how to present and how to use visual tools, respectively. Students should have Data visualization / Infographic skills in order to be able to effectively present the results of data analysis to the user or executive member.

3.3 Focus group discussion Indonesia

This section presents the findings from the Indonesian focus group discussion from 3 different perspectives: ICT industrial sector, student and academics, respectively.

Issue 1: Identification of subject areas related to data science and AI that are most in need in Indonesia (and wider region)		
Industrial Sector Perspective:	Students Perspective:	Academics Perspective:
• Data scientist and AI-Engineer are highly needed. • Candidates should understand Probability theory and Statistics and have ability in Programming (preferable Pyhton). • As AI engineer, the candidates should have abilities and understanding in AI-related knowledge; such as Machine Learning, Neural Networks, Information Retrieval, Natural Language Processing, Computer Vision, Language and Speech Technology, Computational Intelligence • As Data Scientist, candidates should have abilities and understanding in Text mining and Information Retrieval, Machine Learning, Data Modelling and Management. • Start-ups tends to request full-stack capability from the candidates, therefore if the programme can provide wide spectrum of data science and	• Related to curriculum, students tend to follow what was prepared by the University or data science and artificial intelligence programme. • Based on questionnaire, as long as the programme will lead them for becoming data science and artificial intelligence related master, they will follow everything which is necessary and necessity that have been prepared by University or study programme.	• To join the master programme in data science and artificial intelligence, candidates must already be familiar with programming. • For USU&Unsyiah: Candidates need to have knowledge of basic subjects such as: Probability Theory and Statistics, Linear Algebra, Discrete Mathematics. DS related subjects such as; Management Information Systems, Knowledge Representation, Data Structures and Algorithms, Data Modelling and Management. AI related subjects such as; Machine Learning, NLP, Business and Market Analysis, Computer Vision, Ethical, Legal and Social Aspects of AI, and some selective subjects such as; Language and Speech Technology, Robotics, and Virtual Reality and Gaming. • For ITB: knowledge in Data structures and algorithms is

artificial intelligence knowledge, they do not have any objection at all.		pre-requisite for the candidates in order to join the master programme in data science and artificial intelligence. After joining the programme, students learn several subjects which are related to data science and artificial intelligence, such as discrete mathematics, text mining and Information retrieval, probability and statistics, machine learning, NLP, business analytics, computational intelligence, computer vision, knowledge representation, big data, and AI engineering which related to deployment model for large scale environment.

Issue 2: Identification of a set of skills that the curriculum should develop and promote

Industrial Sector Perspective:	Students Perspective:	Academics Perspective:
• Industrial partners also agree to make the curriculum to be 80% about technical skills and 20% about soft skills. • As per priority, they put analytical skills as the first skills that should be developed by the candidate or be developed by curriculum, followed by Empirical methods, Modelling, Teamwork, Written and Oral Communication, Argumentation and Presentation. • If possible, after graduation, they also able to have capability to teach and share knowledge with other people.	• The curriculum should provide every aspect of skill that is needed or required by the industry. • Students tend to believe in the curriculum that will be provided by the University.	• Academics tend to agree to make the curriculum to be 80% about technical skills and 20% about soft skills. • Those skills include: Analytical skills, Modelling, Empirical Methods, Teamwork, Written and Oral Communication, Argumentation and Presentation, Reimplementation of Existing works, and continuous learning in term of comprehension of scientific articles. •

Issue 3: Teaching and learning processes that are appropriate for the curriculum (project-based learning, professional certifications, trainings, practical and industrial projects, workshops, internships, research, theses, etc.)

Industrial Sector Perspective:	Students Perspective:	Academics Perspective:
• Short course/training from the university is one of the alternatives to upgrade the skill, but they tend to see the content of that short course rather than the certificate that will be received. • Internship is one of the ways to recruit the talents, therefore they are very welcome if the internship is a compulsory in this study programme. • Industrial partners also accept students who want to work as part-timer. • Exploration of tools in regards with data science and artificial intelligence can be beneficial for industrial partners.	• The curriculum should provide every aspect of skill that is needed or required by the industry. • Students tend to believe in the curriculum that will be provided by the University. • The university should organize classes which invites industrial partners. They are necessary to get insight as well as to motivate the students.	• Academic staff tend to easily accept any methodologies to be conducted during the programme. • For USU & Unsyiah: they suggested project-based learning for final project or theses writing. An internship is necessary for the students to be aware with the industrial problem, and this internship should be done within 2 – 3 months during the semester break, and it will be great if offer of internship for the students are coming from Industrial partners. • For ITB: they suggested project-based learning, practical-and-industrial projects, and internship as the teaching methodologies for developing the practical and awareness of the real application. As research and theses, as well as joining short course from the university can be used as an alternative to strengthen the understanding in the subjects. ITB also agree for the internship to be done within 2 – 3 months during the semester break, to expose the students with the real-world application.

Issue 4: Required resources, facilities, tools, as well as support that the universities should provide (what we lack or need)

Industrial Sector Perspective:	Students Perspective:	Academics Perspective:
• Industrial partners hope that universities will provide a specific room which is dedicated for discussion of data science and artificial intelligence, room that caters students' discussion, especially if it is related to technical issue.	• Facilities such as high-end workstations, dedicated discussion rooms and laboratories for data science and artificial intelligence, suitable internet connection for downloading big datasets.	• Universities in Indonesia still lacking in certain ways, such as suitable workstations for students for big data analysis, cloud computing resources, private servers for big data, internet connection that is more reliable, research lab that accommodates data science and artificial intelligence programme specifically, and also licenses in regards with software and tools.

Issue 5: Any concerns and opinions regarding internships and job opportunities that the curriculum should provide

Industrial Sector Perspective:	Students Perspective:	Academics Perspective:
• Industrial partners tend to prefer longer period for internship (around 3-6 months). Although it is very rare, but students with longer period of internship do not mean to be performed better in meeting their expectation. In regards with this time period, industry will accept any period that is designed by the University (minimum 2 months).	• For internship, students only follow what the curriculum, if it is 3 months, then they will happily do it. • Job opportunities are not an issue in our Universities, but more access to giant multi-national companies will be better. • Students prefer to become entrepreneur, applying their knowledge after maximum 5 years working for the company.	• Inviting external experts to classes to discuss future career possibilities is considered as very good idea to inspire students.

Conclusions

Based on Industrial needs in Indonesia, Data scientists and AI-engineers are still needed for multiple years ahead by IT-related jobs/companies. Therefore, from the industry point of view, data science and artificial intelligence programmes are beneficial for the industry in Indonesia. Among Universities, we agree to put capabilities in programming as compulsory for the students before joining the data science

and artificial intelligence programme, however discrepancies happen in certain level of knowledge as pre-requisite by students before joining the programme because of each university policy. Detail of the course mapping can be found in Table 1.

Between academics and Industry, we both agree for the curriculum to be 80% for technical skills and 20% for soft skills that can be included within every subject as class-project or tasks. To produce graduates who are more ready for the job, academics and industry also agree to put industrial internship inside the curriculum. Even though Industrial partners prefer longer period for the internship (6 months) but industrial partner agree to follow the period of internship as designed by University (3 months). Besides the internship, Industrial partner also open for the students to participate as part-timer in their office. The industry also encourages students to explore more in regards with the data science and artificial intelligence tools, so the students should follow the current trends in data science and artificial intelligence. Short course in specific domain of data science and artificial intelligence is interesting for industrial partners, as one way to upgrade the skill of the employee. The industrial partners tend to look for the contents of the short course regardless the certification.

The laboratory equipment, suitable internet connection should be prepared and ready to use for the most effective learning. The tools used in teaching and learning should be a tool that is in line with those commonly used in the industry. Students should have the skills to use those tools that can lead to the product development, not just a tool that running or testing as experiments. The course must be built up the collaboration with the industry in applying the real case study for teaching or for students to experiment.

Not only hard skills that students can gain through education, training programmes and internship training but also soft skills for data science and artificial intelligence students is very important. The soft skills that are suggested to be developed including (1) Communication skills, (2) analytical skills/ empirical skills and (3) Presentation skills, both in using visual tools and in preparing content, respectively. Students should have Data visualization / Info graphic skills in order to be able to effectively present the results of data analysis to the user or executive member. (4) Teaching skills, for them to be able to teach their junior/colleagues.

3.4 Focus group discussion Sri Lanka

This section presents the findings from the focus group discussion in Sri Lanka from 3 different perspectives: ICT industrial sector, student and academics, respectively.

Issue 1: Identification of subject areas related to data science and AI that are most in need in Sri Lanka

ICT Industrial Sector Perspective:

- Industry partners strongly believe in the necessity of a well-defined data strategy, which enables the organizations to structure the data centric activities and treat data as an asset rather a commodity. Without such long-term view, organizations find very difficult to focus activities

towards a common goal. Hence, the identification and implementation of suitable data strategy must be included in the curriculum, as per their opinion.

- Emphasis was given on the domain knowledge and the ability to apply data science knowledge within the domain, as it is important for students when they have to work in the industry projects.
- Blockchain was another subject among industry concerns. This topic was described as the next biggest thing after the internet and it has already created lot of excitement within the context.
- All industry participants agreed upon the value of studying data security, as it is important to protect data. Industries were dealing with lot of sensitive information of their customers.
- Laws and ethics in data science was identified as essential to restrict how you handle data. This is very important in the industry, and the employees who wish to join in the fields of data science should have awareness in them. Ex: GBDR (General Data Protection Regulation), For instance, nowadays the customers have a right to forget their data at a given company upon their request.
- Since the technology is continuously changing, the students need to have the knowledge in a wide range of tools and the ability to adapt to new tools quickly. For that reason, the students should essentially have the exposure to different tools like, Hadoop, Spark, Cloudera along with the practice in handling large data sets.
- Real time data handling is widely used in the industry and for that, the tools like Kafka is also useful for complex event analysis and processing.
- Data Governance – This refers to the overall management and caretaking of the data covering its usability, integrity (i.e. making sure the data is of a good quality, that you know where it has come from and that you have the right to use it as you need) and security. As discussed, this can be inserted in the curriculum as a part of data laws and ethics.
- Data Science Methodology – This defines the essential steps that must be followed towards a comprehensive study. The methodology steps can be; Problem to Approach – Business understanding, Data preparation – Cleaning & transforming data, Data understanding – Can the available data answer the question at hand?, Analytic approach – Statistically/machine learning driven?, Modelling – In which way we can use the data to answer the given question, Evaluation – Does the model answer the question, or does it need adjustments?, Deployment – Making the solution to be used by the stakeholders, Feedback – Feedback from the users and refine the model as needed.
- Lack of Mathematics knowledge in the employees make difficulties in understanding the domain problems related to machine learning. Mathematical optimization, linear algebra and multivariate calculus are also important for a master programme in this scenario.
- Subjects related to statistics and mathematics such as Statistical Data Analysis, Statistical modelling and statistical learning theory, Mathematical Optimization were proposed.
- In addition to theory, the applications of Deep Learning, Machine Learning and Data Science were mentioned as necessary for the course.

Student Perspective:

- Students pointed out the importance of learning Advanced Database Management Technologies such as NoSQL databases (e.g., HBase, Cassandra) for storing and processing big data in distributed environments.
- Containerization has risen up in recent years as a ready to use solution to work around platform differences across a variety of applications from single user development environments to highly

distributed production environments (e.g. Google Cloud, Docker). Students were interested in studying trendy topic like this.

- More practical knowledge is important for doing a research successfully. May be students are not studying what they actually want for their research implementation. For that, more practical sessions on a variety of tools in DS&AI is advantageous for students.
- Vector calculation was also suggested as important for data analysis.
- Students also proposed Block chain programing, as it would be a latest topic in DS&AI in near future.
- Students further suggested financial data analytics with AI as a part of regression analysis which would be useful in analysing financial data.
- In addition, topic such as Expert systems in AI, Deep learning for AI and Mahout for large scale Machine Learning were also proposed as essential for in depth knowledge in DS&AI.

Academics Perspective:

- Not having enough data for students to practice data science concepts and to learn the techniques is a problem. Companies might also be reluctant to share data due to their sensitivity. Hence, we have to encourage students to use open data sets that are now largely available.
- Advanced network technologies, High Performance Computing and Distributed Computing were identified as important subjects for students to have the knowledge used when working with large data sets. At least, the basics of them were proposed to include in the curriculum.
- To understand advanced statistical tools, enough Mathematics background is essential. Nonlinear programming and graph theory were identified as significant for neural networks.
- Matrix factorization was also proposed as important for big data analysis.
- Basic financial knowledge and economics with foundation in financial knowledge was identified as useful to interpret values and matrixes.
- Business management to give emphasis on the importance of business and understand the business perspective in terms of DS&AI was suggested for the curriculum.
- Scientific Report Writing should be developed for students to be able to effectively describe the data appropriately.
- Operating hardware components that are related to programing like GPU was found useful for DS&AI.
- Big data processing and large scale machine learning and Deep Learning were suggested as core topics that should be included in the course. Under the course, following topics such as; Distributed databases (e.g. HBase), Data Transfer (flume, sqoop), Large Scale machine learning (Mahout, MLib), Real time data analytics (kafka, storm), Security – (zookeeper) and Administration – (Ambari) were proposed.
- Introduction to Artificial Intelligence should be included to provide foundation knowledge in AI. Advanced AI with topics in the areas of knowledge representation and general artificial intelligence was also suggested.

Issue 2: Identification of a set of skills that the curriculum should develop and promote

ICT Industrial Sector Perspective:

- One essential skill which was pointed out here was Presentation skills. In a corporate setup, how you tell your story according to the gained insight from data to a non-data savvy business leader is critical. No matter how important your insight is, if the final communication fails everything starts falling around it. So, building your story with right visualization in a clear concise way is a critical skill to develop.
- Documenting data science results and implementations in a standard way was proposed as very important for better understanding and to avoid unnecessary changes.
- Having good analytical skills to understand the business problem, being innovative, meeting deadlines and being productive, project management and effective communication skills in scientific communication were also suggested as required skills for the industry.

Student Perspective:

- M.Sc candidate should be able to do effective presentations and communication. In addition, team working, scientifc writing skills and and the self-education skills are important.

Academics Perspective:

- Ability to use a version control system
- Ability to adapt into changing technology

Issue 3: Teaching and learning processes that are appropriate for the curriculum (project-based learning, professional certifications, trainings, practical and industrial projects, workshops, internships, research, theses, etc.)

ICT Industrial Sector Perspective:

- Hackathon related data science
- Github open source projects for every student.
- Project based learning always gives an opportunity to practice and replicate end-to-end (problem approach -> deployment + feedback) data science methodology.
- Professional certification should be awarded and ideally it could be in a digital form so participants can add them into their on-line profile (saves paper too).
- Suggest having cross department collaboration to reinforce researches where Data Science students can provide the data specific insights to the research.

Student Perspective:

- All the suggested teaching and learning processes given here can be used.
- If the students can also get the professional courses, that would be very helpful in their studies as well as in getting to industrial opportunities.

Academics Perspective:

- A compulsory module for research should be offered in the course for doing extensive individual studies on a researchable topic. The industry companies were invited to share research areas with students and supervise them accordingly. An evaluation criterion should be set to publish at least one research paper.
- Offering students a compulsory mini project to get the exposure for different technological environments common in the industry, along with assignments shared by the industry.

- In addition to class room teaching, project based learning should be used.
- Sort term training courses should be available for students as well.
- Problem based learning

Issue 4: Required resources, facilities, tools, as well as support that the universities should provide (what we lack or need)

ICT Industrial Sector Perspective:

- If we can initiate a discussion with Google, Amazon, Microsoft, IBM, and other large companies to partner up to utilize their cloud infrastructure to the participating students at a concession rate it will ease up lot of infrastructural burden to be provide to students. We never be able to catch these tech giants on these aspects. This would add a great value to the course content itself.
- Use machine learning as a service. Ex: AWS(Amazon Web Services) – SageMaker

Student Perspective:

- Learning Management System to make different aspects of the course online. (lecture material, evaluation, etc)
- It would be better if there are more poweful machines for DS&AI projet implementations.

Academics Perspective:

- Facilities will be required to get high performance computing power if necessary.
- Data repositories and clean data are much needed.
- The university has a Data Science Unit (DSU) that is implementing a data sharing platform, where the industries can share their preferred data sets, and ultimately this would be beneficial for researchers. It consists of a computer cluster using 3 powerful machines with the following specifications[4]:

Model	Processor(s)	Memory	Hard drive capacity	GPU	GPU Memory
Dell Poweredge R740xd	2 x Intel Xeon Gold 6130 2.1G, 16C/32T	128 GB	3 TB	Nvidia P40	24 GB
Dell XPS 8920	3.6 GHz Intel Core i7-7700 Quad-Core.	32 GB	2 TB	Nvidia GeForce GTX 960	8 GB
Dell XPS 8910	3.4 GHz Intel Core i7-6700 Quad-Core	16 GB	2 TB	Nvidia GeForce GTX 660	2 GB

We have set up a Hortonworks cluster for Big Data analytics. Moreover, we tested the applicability of Hadoop tools such as Pig, Hive, HBase, Sqoop, Flume, and Mahout, and prepared documentation explaining students/staff how to use those tools on our cluster. In order to

[4] Numbers mentioned in the table may be slightly different.

distribute computing resources between multiple users to run their programs concurrently, we are using SLURM as the scheduler and resource manager. We have also prepared documentation for how to run programs using SLURM. All the documents are available in our Data Science Unit website.

The cluster supports effective utilization of resources for deep learning projects using python frameworks such as Pytorch, Keras, Tensorflow, and Theano. Moreover, we provide additional support for the users who remotely run R programs using Rstudio Server that is running on our cluster. Recently, we conducted a deep learning lab session for the M.Sc. in Computer Science 2018 students, where around 50 students remotely accessed our cluster and ran their programs.

We assume our resources would be adequate to start an MSc in Data Science and AI course. In the future, we may purchase additional machines and integrate them with our cluster. We hope we may not need to rely on a cloud platform. If need increases, we may consider using cloud infrastructure along with our cluster. We keep updating our cluster to incorporate modern tools that are necessary for the MSc in Data Science and AI.

Issue 5: Any concerns and opinions regarding internships and job opportunities that the curriculum should provide

ICT Industrial Sector Perspective:

- If the MSc student is already working in another company, providing projects for them with real data would be challenging.
- From the very beginning we must develop a strong communication strategy on the program so, stakeholders (including the industry) see the benefits right away. Academic and business sense needs to be properly intertwined in the overall message.
- Deep Learning use cases like face recognition can be done without data sharing issues.

Academics Perspective:

- Trainig in the idustry should be a compulsory module and its duration has to be a minimum of six months.
- Implement a proper NDA and MOU with the industry partners.

The focus group discussed an additional issue that is worth reporting:

Issue 6: Identification of short term professional training courses in data science and artificial intelligence

ICT Industrial Sector Perspective:

- When a company sends someone for a short term course, they expect quick results in return. A course on Exploratory Data Analysis was suggested to provide competencies in identifying data patterns, visualization and interpretation. Basic mathematics and statistics knowledge would be pre-requisites for interpretation of results. Delivering of the course must need interactive sessions, not just listening mode lectures. Minimum required duration was decided as seven days which should facilitate:

- Language: R or Python (Python is most widely used in the industry, due to distributed environments)
- Data preparation/ Data cleaning: Pandas/ dplyr (Domain knowledge is also important)
- Visualization: ggplot2, Seaborn, Matplotlib, Power BI, Tableau
- Data analytics:SAS, SPSS

- Courses that can support to gain the abilities in working on industry grade cloud infrastructure (IBM Watson, Microsoft Azure, etc.).
- Courses that facilitate the basic practice of using tools such as Github
- It would be ideal if the courses can cater to three types of audiences, such as, employees with software developing knowledge, employees with statistics and mathematics backgrounds and employees from management and physical sciences streams.
- A proper evaluation mechanism should also be designed to particular course, which would be credible to the company. If the exams are automated and available online, that would be ideal for professionals.
- Courses on Electronic Data Interchange would be useful as in most cases we assume that data is in the big data system, but actually we need to get a lot of effort to transfer data to there.
- Short courses for Python basics and standards, as well as for, TensorFlow, caffe, paddle, Keras, Pytorch and deep learning frameworks were mentioned.
- Object oriented programing basics courses for statistics and mathematics students.

Student Perspective:

- Providing detailed knowledge in a set of Hadoop ecosystem that are widely used in local industries (e.g., Apache Kafka for real time data processing).
- Course on Advanced Machine Learning in Python was also proposed, with the tools such as scikit-learn, Pytorch and Vowpal Wabbit. Content of the course was proposed as, Traditional machine learning algorithms, artificial neural networks, deep neural networks, convolutional neural networks, long-short term memory networks, word embedding, ELMo, Transformer networks and BERT.
- Courses in Big data and Hadoop Ecosystem was suggested along with the knowledge on the topics such as; Distributed databases (e.g. HBase), Large Scale machine learning (Mahout, MLib), Data Transfer (flume, sqoop), Real time data analytics (kafka, storm), Security (zookeeper) and Administration (Ambari)

Academics Perspective:

- If a particular course can provide elective and supportive subjects as well, then the participants will be able customize the course according to their competence level.
- A short course on how to use Python in a developer environment of a real data science project.

Appendix A: Focus Group Discussions

Focus Group Thailand

A focus group discussion was conducted jointly by the three partner universities in Thailand: (AIT, KKU and WU), on Thursday 1st August 2019, during 13:30-15:30 via online (zoom teleconference). The discussion group composed of 4 representatives from ICT industrial sector, 3 students, and 8 academics, as shown below.

Representatives from ICT industrial sector:

1. Mr. Chinnawat Devahastin Na Ayudhya
   Senior Analyst, Data Analytics Group, Information Technology Group
   Bank of Thailand (BOT), Sector: Banking, Size: 250 or more
2. Mr. Chana Supatsorn
   Deputy CEO, Intelligist Company Limited,
   Sector: Big Data Platform Provider, Size: 20-50
3. Dr. Kan Ouivirach
   Software Engineer Manager
   Pronto Group Ltd.
   Sector: Software House, Size: 100-150
4. Mr.Tanin Sammanee
   CEO, Deepscope, FinTech Startup Company,
   Sector: Startup and FinTech, Size:

Students:

1. Mr. Akraradet Sinsamersuk, Computer Science master programme, AIT
2. Miss Sasipim Prasartsri, IT master programme, KKU
3. Miss Thohirah Husaini, Software Engineering master programme, WU

Academics:

1. Dr. Chutiporn Anutariya, ICT Department, AIT[5]
2. Dr. Wararat Songpan, CS Department, KKU
3. Dr.Chitsutha Soomlek, CS Department, KKU
4. Dr. Suratsavadee Korkua, Department of Electrical Engineering, WU
5. Dr. Chirawat Wattanapanich, Department of Computer Engineering, WU
6. Dr. Thimaporn Phetkaew, Department of Software Engineering, WU
7. Dr. Putthiporn Thanathamathee, Department of Software Engineering, WU
8. Dr. Siraporn Sakphrom, Department of Electrical Engineering, WU

[5] Also acting as a focus group moderator.

Focus Group Indonesia

A focus group discussion was conducted jointly by the three partner universities in Indonesia (ITB, USU and Unsyiah), on Thursday 20st August 2019, during 15:00–16:30 via online (zoom teleconference). The discussion group composed of 7 representatives from ICT industrial sector, 4 students, and 9 academics, as shown below.

Representatives from ICT industrial sector:

1. Ali Septiandri – airyrooms
2. Arie Sutiono – airyrooms
3. Samsu – airyrooms
4. Muhammad Ghifary – Bukalapak
5. Juanda Lokman – ions tech
6. Budi Prasetyo – Mandike
7. Teguh Eko Budiarto – Prosa

Students:

1. Denny syaputra – Unsyiah
2. Teuku putera – Unsyiah
3. Adine Belinda – USU
4. Haslin Lawalata – USU

Academics:

1. Nugraha P. Utama – ITB[6]
2. Ayu Purwarianti – ITB
3. Masayu Leylia Khodra – ITB
4. Taufik Abidin – Unsyiah
5. Opim Salim Sitompul – USU
6. Emerson Sinulingga – USU
7. Erna Budhiarti Nababan – USU
8. Maya Silvi – USU
9. Elvi Zamzami – USU

Focus Group Sri Lanka

A focus group discussion was conducted by the University of Peradeniya (UoP) and the University of Sri Jayewardenepura, Sri Lanka on Saturday 10th August 2019, during 9.00 – 12.30, with the participation of 4 ICT companies, 3 students and 13 academics, as shown below.

Representatives from ICT industrial sector:

1. Mr. Pradeep Indrajith
   Principal Data Analyst – Business Operations

[6] Also acting as a focus group moderator.

IFS Sri Lanka
Sector: Software vendor, Size: over 700

2. Mr. Shanika Amarasoma
   Principal Architect-Data Engineering
   Axiata Digital Labs (Pvt) Ltd
   Sector: innovative software service provider, Size: over 400
3. Ms. Nuzla Ismail
   Data Scientist
   Axiata Digital Labs (Pvt) Ltd
4. Mr. Hansa Perera
   Data Scientist and Machine Learning Researcher
   Zone 24x7 (Pvt) Ltd
   Sector: technology solutions provider, Size: 201-500

Students:

1. Mr. Kayanan Manickavasagar, PhD candidate in Statistics, UoP
2. Mr. Jarashanth Selvarajah, M.Phil. candidate in Computer Science, UoP
3. Ms. Amirthavarshini Mahadevan, M.Sc. candidate in Computer Science, UoP

Academics:

1. Prof. Saluka R. Kodituwakku – Professor in Statistics & Computer Science, UoP
2. Prof. Pushpakanthie Wijekoon – Senior Professor in Statistics, UoP
3. Prof. Roshan D. Yapa – Professor in Statistics & Computer Science, UoP
4. Dr. Chitraka Wickramarachchi, Dept. of Statistics, University of Sri Jayewardenepura
5. Dr. Ruwan D. Nawarathna, Dept. of Statistics and Computer Science, UoP
6. Dr. Amalka Pinidiyaarachchi, Dept. of Statistics and Computer Science, UoP
7. Dr. Hakim Usoof, Dept. of Statistics and Computer Science, UoP
8. Dr. Lakshika Nawarathna, Dept. of Statistics and Computer Science, UoP
9. Dr. Hemalika Abeysundara, Dept. of Statistics and Computer Science, UoP
10. Mr. Prabhath Gunathilake, Dept. of Statistics and Computer Science, UoP
11. Dr. Erunika O. Dayartna, Dept. of Statistics and Computer Science, UoP
12. Dr. Shameen Jinadasa, Director , International Affairs Office, UoP
13. Dr. Nanda Gunawardena, Former Director , International Affairs Office, UoP