Development of Data-Driven Curriculum Courses in the Bachelor of Engineering Degree Program

M. Rantonen
JAMK University of Applied Sciences

M. Karjalainen
JAMK University of Applied Sciences

The rapid technological development is typical for the IT industry, and it poses a challenge for the curricula to stay up to date. It can also be problematic to bring new technologies into the curriculum so that the contents and views of the technologies have sufficient maturity. The aim of this paper is to illustrate how the competences of the developed Data analytics (DA) and Artificial Intelligence (AI) curriculum in a higher education organization map to the requirements of data-driven real-life work. In spring 2017, the Institute of Information Technology at Jamk University of Applied Sciences started to develop its competence in DA and AI. Based on the experience of the research team, each use case is divided into substances, and the substances are categorised as courses which were formed for the 30-credit study module. After this, the course content is generated from the learning outcomes and they are reflected on the data-driven process model and the competencies defined based on the model phases.

Keywords: artificial intelligence, data analytics, curriculum courses, big data, emerging technology, virtual courses, machine learning, deep learning

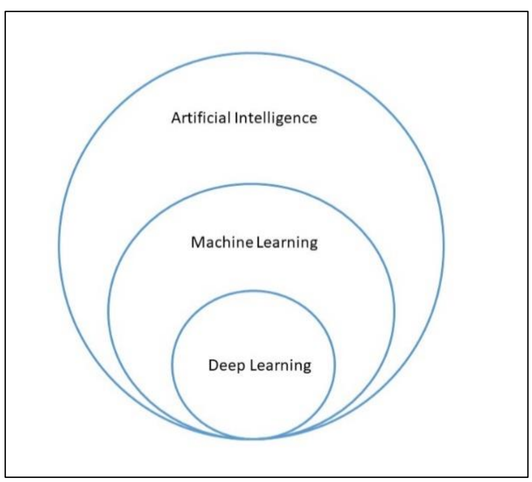
INTRODUCTION

In spring 2017, the Institute of Information Technology at Jamk University of Applied Sciences started to develop its competence in DA and AI. During these three years, various data-driven based methods (such as data analytics, machine learning, deep learning.) were developed for the different data sets to achieve broad knowledge from the different domain. The lack of good and well-organized university level education material and knowledge of the previously mentioned technologies were the major challenges during the development process. Jamk's expertise is developed by a research team consisting of staff and students, whose approach can be described as learning by doing, i.e., handson. Almost 30 students participated in the work of the research group in 2017-2020, completing a practical training of 30 credits, which is part of ICT Engineering Bachelor’s degree program. Based on the experience of the research team, each use case is divided into substances and the substances are categorized as courses which were formed for the 30-credits study module in question. The design of the contents of these courses is based on our experience of individual, unique and solved data-driven problems using the methods mentioned above. The publication also presents the developed courses and their learning outcomes. After this, the course content is generated from the learning outcomes and reflected on the data-driven process model and the competencies defined based on the model phases. Furthermore, the course contents of the study courses are mirrored to the actual data-driven job titles used in working life. Due to the difference between project management for data-driven projects and traditional software projects, the job descriptions and corresponding job titles also vary drastically. Thus, several new job titles have emerged in the sector in recent years, but they are already becoming well established.

FIELD OF ARTIFICIAL INTELLIGENCE

At present the data-driven project (known also as a data-based) can include different types of projects such a data mining, data science, data-analytics (DA) and artificial intelligence (AI) projects. Further, AI can be divided into machine learning (ML) and deep learning (DL) (Goodfellow, 2016), see Figure 1. ML is a collection of algorithms and techniques used to design systems that learn from data (Alpaydin, 2021). Respectively, DL is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit, far from matching its ability—allowing it to "learn" from large amounts of data (Kelleher, 2019). Data are key elements during the algorithm development and due to these, systems are able to perform predictions or deduce from the supplied data. ML algorithms can be categorized as supervised learning algorithms and unsupervised learning algorithms. ML algorithms have a strong mathematical and statistical basis without domain knowledge (Lee, 2019). Supervised learning algorithms are trained with labeled data i.e., data composed of examples of the desired answers. For instance, a model that identifies anomaly use of credit cards would be trained from a dataset with labeled data points of known anomaly and valid use. Most ML algorithms are supervised, for example classification and regression type algorithms. Classification is identifying to which set of categories a new observation belongs based on the set of training data in the observed categories. The most used example is an email spam filter in which the filter classifies the emails as *spam* or *not spam*. The regression helps in forecasting or predicting the future by estimating the relationship between variables, which means that the output is a continuous output variable. Correspondingly, unsupervised learning algorithms are used on data with no labels, and the goal is to find relationships in the data. Clustering belongs to unsupervised learning algorithms and its goal is to group similar data points into intuitive groups.

FIGURE 1
A VENN DIAGRAM SHOWING THE RELATIONSHIP BETWEEN AI TERMINOLOGY



All the above-mentioned data-driven projects are based on data, which can be used either to develop models or to find new information from the data. The amount of data exponentially grows and with the fast development of technology (van der Aalst, 2016), and the data utilization has recently become an upward trend. Higher education has to respond to changes in the real-world working life and therefore the development of new curriculum courses is essential. This paper describes the structure of the developed data-driven curriculum in a higher education organization and the ways how the achieved course competencies reflect to the requirements of data-driven real-life work.

One focus of the DL is to create a large neural network model that is capable of making accurate data-driven decisions. The idea behind the DL network is a mathematical model that is (loosely) inspired by the structure of the human brain (Kelleher, 2019). The abstraction of human brains, neural network, is called an artificial neural network (ANN) consisting of a network of simple information processing units called neurons as in the human brain. The power of neural networks is to model complex relationships without the complex mathematical models, but rather it emerges from the interactions between a large set of simple neurons. All ANNs are not the DL networks because DL networks have many hidden layers of neurons. The minimum number of hidden layers necessary to be considered as deep is two. However, most DL networks have many more than two hidden layers. The depth of a network is measured in terms of the number of hidden layers plus the output layer (Kelleher, 2019).

Due to the rapid proliferation of data-driven projects, there has been a major change in the job market in terms of job titles and job descriptions. Thus, new jobs need various range of well-knows skills and as well as new skills. Many data-driven project models have been published long ago and are referred to as data mining or data warehousing process models. The data-based project differs significantly from the traditional software development project. However, several different methods have been used in recent years, which are directly developed from software development projects. Various process models and frameworks such as CRISP-DM (CRoss Industry Standard Process for Data Mining) (Shearer, 2000), Knowledge Discovery in Database (KKD) (van der Aalst, 2016), (Feyyad, 1996), Team Data Science Process TDSP) (What is the Team Data Science Process?, 2022), Foundation Methodology for Data Science (FMDS) (IBM, 2022) describe how to execute a data-driven project. Correspondingly, Scrum (Pries, 2010), (Pichler, 2010) and Kanban (Brechner, 2015) are pure software project management methods but also customized versions for the data-driven projects are developed. Based on the data-driven project models, the different phases can be categorized into the real-life jobs based on which the necessary competencies can be derived accordingly.

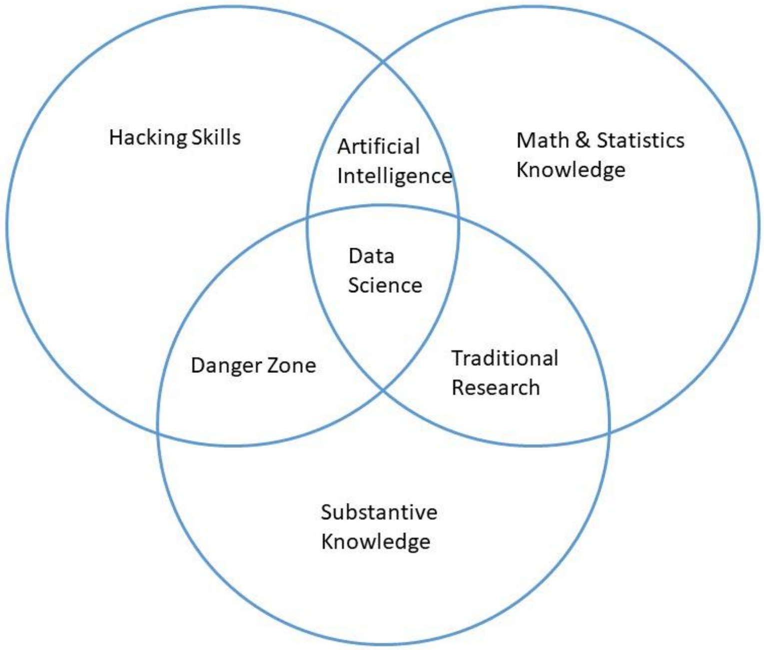
Defining the data science/analysis skills is the split between substance and methodology which are ambiguous and unclear in how to distinguish among hackers, statisticians, subject matter experts, their overlaps and where data science fits. Therefore, a fully competent data scientist/analyst needs many different types of very extensive skill sets. Jamk's approach to learning has been hands on training, which has proven to be a very good solution. Knowledge has been gained precisely by developing different types of applications and use cases. The application of DA and AI differs significantly from the traditional use of program libraries in that the application requires a profound knowledge about the data and domain. Knowledge of theory alone does not suffice; the developed DA and AI methods must be selected carefully and tested in practice. In addition to the practical hands-on work of the research team, it was noticed that the prolific knowledge can be transformed into courses that can be used in the future for university-level teaching. Due to the requirements of skills, the courses and contents have been derived from the developed DA and AI applications and use cases.

DATA SCIENCE PROCESS MODELS

Methodology

Data-driven projects are emerging from the intersection of several up-and-coming technologies, but nowadays the technology is not the critical point for a successful project. DA and AI projects are a relatively new industry, albeit the fact that its components have been around for a long time. Thus, the success of a data-driven project depends on the effective execution and management of the project. Venn diagram in Figure 2 is used to understand that data science and AI are a combination of several disciplines. In this Venn diagram, the three components are hacking skills, math and statistics knowledge, and substantive expertise (The Data Science Venn Diagram, 2021), (The Essential Data Science Venn Diagram, 2022).

FIGURE 2
A VENN DIAGRAM SHOWING THE RELATIONSHIP BETWEEN DATA ANALYTICS/SCIENCE AND OTHER SKILLS



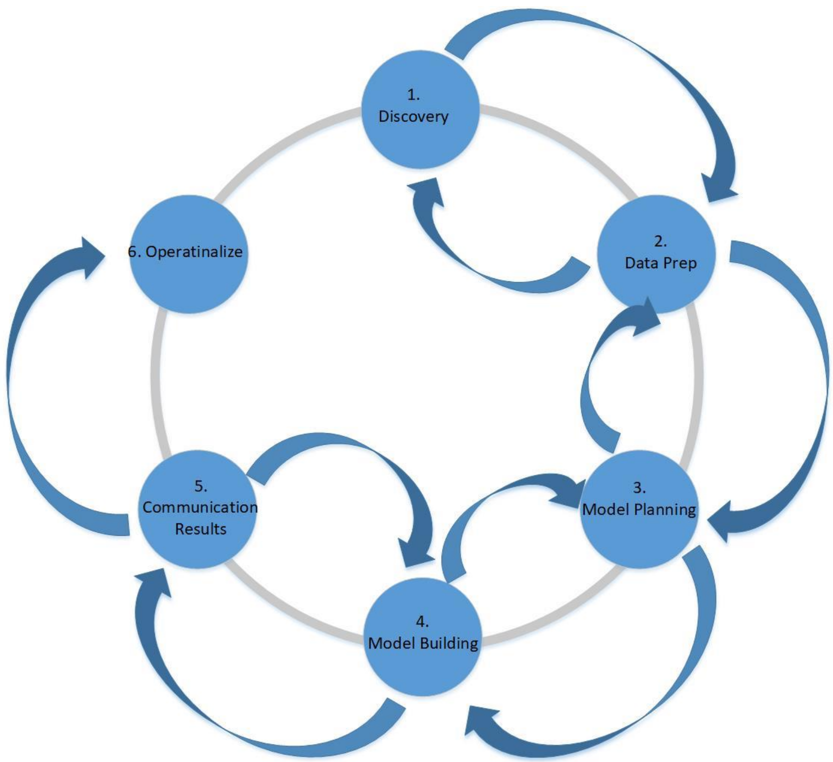
Data-Driven Project Lifecycle

In generally, the data-driven project can be divided into 6 different phases (EMC, 2015):

1. Discovery
2. Data Preparation
3. Model Planning
4. Model Building
5. Model Communication Results
6. Operationalisation

The project work can take place in several phases at simultaneously and the movement can be either forward or backward. This iterative depiction of the lifecycle is intended to portray a real project more closely, in which aspects of the project move forward and may return to earlier stages as new information is available, or team members learn more about various stages of the project or domain. This enables participants to move iteratively through the process and drive toward operational and successful project work (EMC, 2015).

FIGURE 3
OVERVIEW OF DATA ANALYTICS LIFECYCLE

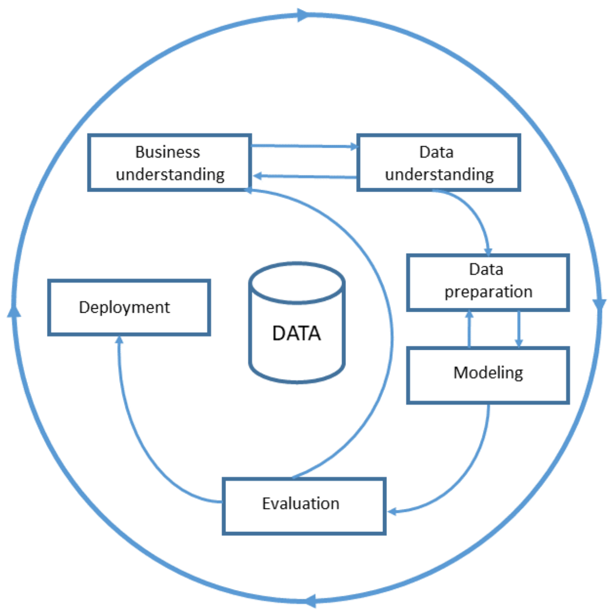


It was previously mentioned that CRISM-DM (Shearer, 2000) is the most commonly used data-driven process model. In general, all data-driven process/project models consist of four common iterative stages including problem definition/ formulation, data gathering, data modelling, and data product development. The goal of the data-driven project could be to enhance or improve decision-making process by providing data-driven predictions. This requires principles, processes, and techniques to understand the problem through an automated evaluation and data analysis (Foroughi, 2018). Whitaker (2019) describes how data science can be used to generate hypotheses, design experiments, perform experiments, and analyse data. Furthermore, Whitaker (2019) present a vision for how data science techniques will become an integral part of the laboratory of the future.

CRISM-DM

CRISP-DM (Shearer, 2000) is one of the most commonly used data-driven process models and therefore it is used in this publication to describe different phases on a more detailed level. The phases of CRISP-DM are used in the mapping of real-life jobs and curriculum courses. CRISM-DM is an open standard process model that describes common approaches used in data mining, DA, ML and DL projects. In other words, CRISP-DM model standardizes the data mining projects to achieve better results from data mining and to encourage best practices. CRISP-DM breaks down the lifecycle of a data mining project into six phases which help organisations to understand the data mining process and provide a road map to follow while planning and carrying out a data mining project, see Figure 4. The arrows indicate the most important and frequent dependencies between the phases, while the outer circle symbolizes the iterative nature of data mining itself, which illustrates that the process can be iterative and can trigger new, often more focused business questions. CRISP-CM has been used to map special skills and understanding using the developed framework to a specific curriculum course (Rantonen, 2020).

FIGURE 4
CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING



Shearer (2000) describes the detailed phase description of the data mining process as follows:

- **Business understanding** is the most important phase of any data mining project because the understanding the project objectives from a business perspective is very critical. Furthermore, converting this knowledge into a data mining problem definition, and then developing a draft plan designed to achieve the objectives helps to understand which data should later be analyzed, and how, which is critical for data miners to fully understand the business case. The business understanding phase contains the following steps: determining business objectives, assessing the situation, determining the data mining goals, and producing the project plan.
- **Data understanding** starts with an initial data collection and the familiarization with the data which contains identification of data quality problems, discover initial insights into the data, or detecting interesting subsets to form hypotheses about hidden information. The data understanding phase involves four steps: the collection of initial data, the description of data, the exploration of data, and the verification of data quality.
- **Data preparation** includes all activities to construct the final data set or the data that will be fed into the modelling tool from the initial raw data, i.e., the selection of appropriate table, record, and attribute, as well as transformation and cleaning of data for modelling tools. The five steps in data preparation are the selection of data, the cleaning of data, the construction of data, the integration of data, and the formatting of data.
- **Modelling** techniques are selected and applied, and their parameters are tuned to optimal values. Typically, there can be several techniques for the same data mining problem type, and some techniques have specific requirements for the form of data. Therefore, stepping back to the data preparation phase may be necessary. Modelling can be further divided into the selection of the modelling technique, the generation of test design, the creation of models, and the assessment of models.
- **Evaluation** of the model and the review of the model's implementation have to be certain and proper to achieve the business objectives. It is critical to determine if some important business issue has not been sufficiently considered. The key steps here are the evaluation of results, the process review, and the determination of next steps.

- **Deployment** is an end phase of a project, in which the knowledge must be organized and presented in a way that is suitable for the customer so that they can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. The key steps here are the plan deployment, plan monitoring and maintenance, the production of the final report, and the review of the project.

Data Science Project Member

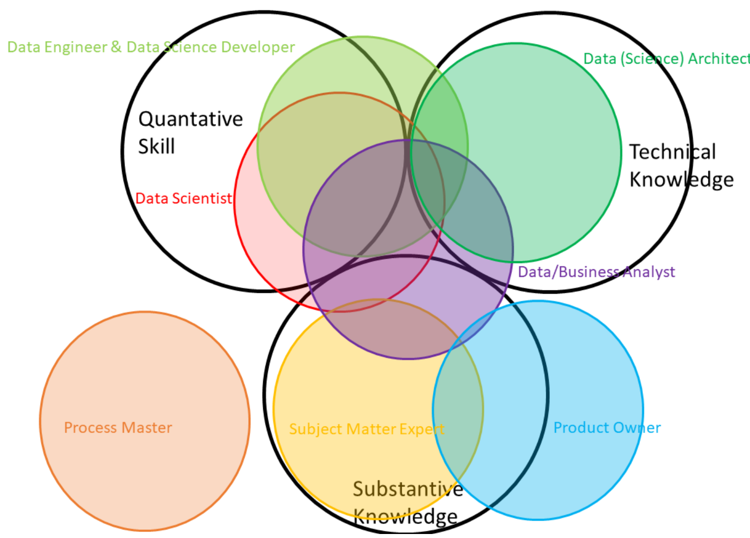
Data-driven based project members can be divided into key roles to consider when one is building and leading a data science team (Grady, 2017). In this publication, eight most commonly used key roles are presented and described (The Essential Data Science Venn Diagram., 2022). These are not in any specific order, as their importance might vary from one project to another or from one organization to another. Furthermore, not all roles are required to be performed by specific or different people. Hence, for a smaller project one person might fill the role of the data scientist as well as the data engineer.

- **Data Scientist** is someone who makes value out of data. Data scientists help companies to interpret and manage data and solve complex problems using expertise in a variety of data sources. Data scientists have a foundation in computer science, modelling, statistics, analytics, and mathematics. Furthermore, a data scientist must have enough business domain expertise to interpret and translate business goals into e.g., data-based deliverables such as prediction engines, pattern detection analysis, and optimization algorithms. Data scientists find and interpret rich data sources, merge data sources, create visualizations, and use DA, ML and DL to build models that aid in creating actionable insight from the data. They know the end-to-end process of data exploration and can present and communicate data insights and findings to a range of team members.
- **Data Engineer** generally transforms data into a useful format for an analysis. More specifically, the data engineer makes the appropriate data accessible and available for data science efforts such as collecting, storing, and processing and cleaning huge amounts of data that can be analysed to gather insights and support decision making activities. Data engineers are vital members of any data analytics team due to their responsibility for building algorithms to help to give an easier access to raw data. The role of the data engineer will vary depending on the type of company such as generalist, pipeline centric of database-centric roles. Data engineers are equally important as data scientists, but they tend to be less visible because they tend to be further from the end product of the analysis.
- **Data (Science) Architect** is considered as a relatively new role in data-based projects that businesses should consider. A Data (science) architect is a job role that is a mix between a data scientist and a data engineer. Data (science) architects design and maintain the architecture of data science applications and facilities. In other words, this role creates and manages relevant data models, data storage systems, processes workflows and data pipeline.
- **Data Science Developer** or **Machine Learning Engineer** designs, develops, and codes large data analytics applications or ML algorithms to support scientific or enterprise/business processes. This role enables models to be deployed and requires some expertise in data science, as well as knowledge of how to effectively develop software applications.
- **Product Owner or Manager** is the central point of product leadership and responsible to decides which features and functionality are built and prioritised in order to build them and what aspects of them to observe and analyse. The product owner also prioritizes tasks ensuring that each work item is clearly defined from a business context and that the upcoming work and priorities of the team are visible and transparent. Furthermore, the product owner must agree that the tasks in the done column are actually done. In short, the product owner represents all the stakeholders in the project.

- **Data/Business Analyst** will take data and figure out a variety of issues depending on the business, for example how to price new materials, how to reduce transportation costs, or how to deal with issues that cost the company. They analyse a large variety of data to extract information about system, service, or organization performance and present them in usable/actionable form. They also shape and interpret a problem for a more suitable form to the data scientist to explore.
- **Process Master** The process master acts as a coach and facilitator, who removes barriers and helps every team member to understand and embrace the project values, principles, and practices to aid the organization in obtaining exceptional results.
- **Subject Matter Expert** is an expert with extensive knowledge of how to apply the analytics or AI within a specific organisational context. This role is accountable to ensure the desired insights and ideas are actionable.

Figure 5 summarized the roles and their competences in the general level mapping to the categorizes: Quantitative Skill, Technical Knowledge, and Substantive Knowledge (The Essential Data Science Venn Diagram, 2022). The job titles here are the most used job titles in the real-working life.

FIGURE 5
SUMMARY OF ROLES AND COMPETENCES



CURRICULUM COURSES

The content of the courses is designed based on hermeneutic circular method (Dickens, 1977). The hermeneutic circular method is particularly suitable for the development of curricula for emerging technologies. While the technologies are still non-mature, there are no well-established frameworks that structures, for example, the needed skills or knowledge. In general, especially in the field of IT technologies, the development of the technologies is so rapid that the timeliness of curricula must be ensured with a continuous hermeneutic process. The hermeneutic circle was formed by the research team that gained the experience by the DA and AI research conducted in the past three years. The course table consists of seven different virtual courses. All courses contain a great number of practical exercises, as is customary in Bachelor of Engineering type of education. Introductory and theoretical types of courses are not available, only the hands-on versions. During the curriculum development process, the staff and students co-operate as a team and develop the ability to carry out concrete real-world DA and AI use cases. The emphasis has been placed on the practical application of open-source AI and DA products, either based on available open data or on a real-life business problem. The application of AI has been facilitated by the introduction of AI libraries of various commercial operators, perhaps the best known is Google's Tensorflow (An end-to-end open source machine learning platform, 2022). In DA, the Python programming language is selected due to its well-known capability in conducting scientific calculations. The diversity of the different types of data and use cases provides a deep insight into the structure and development of the curriculum. The ability and knowledge in DA and AI are well documented and refined to 30-ECTS education and a part of the curriculum in ICT engineering education.

The curriculum consists of the following courses:

- Data Pre-processing, 3 ETCS (European Credit Transfer and Accumulation System)
- Big Data Environments, 5 ETCS
- Data Analysis, 4 ETCS
- Machine Learning, 5 ETCS
- Deep Learning, 5 ETCS
- Project Work on Data Analytics, 4 ETCS
- Project Work on Artificial Intelligence, 4 ETCS

The courses and the contents are derived from field studies. Each field study is from a different domain and each of them has their own challenges. The challenges are mapped using the Framework for course development to the specific course (Rantonen, 2020).

Data Pre-Processing

Learning Outcomes of Data Pre-Processing

The student understands the process of data analytics and its challenges. The student recognizes various data formats, the most common interface solutions and tools and methods used in data processing. The student is able to apply the methods needed in data pre-processing (JAMK, Curriculum of Data Pre-processing course, 2022).

Course Contents

- Various IoT sources/formats, JSON, APIs, queries from the SQL tables
- Types of variables
- Data pre-processing before bringing it into analysis program
- Data pre-processing in Pandas (basics of Pandas and DataFrames)
- Connecting various data sources
- Data encoding

Big Data Environments

Learning Outcomes of Big Data Environments Course

The student understands the significance of data masses in digitalizing operational environments and the challenges and requirements caused by huge amounts of data. The student knows the most commonly used big data systems and their functionalities. The student is able to design a system based on requirement specifications and implement it (JAMK, Curriculum of Big Data Environments course, 2022).

Course Contents

- Big Data as phenomenon
- Big Data applications
- Data pipeline and its stages
- Understanding of the tools applicable to implementation of various stages of pipeline
- Various data processing methods and their implementation (batch and stream)
- Ability to utilize applicable tools to gather, save, process and visualize data

Data Analysis

Learning Outcomes of Data Analysis Course

The student understands the significance of data analytics in the digitalizing operational environment. The student knows the most commonly used methods of data analytics as well as how to apply them in practice to existing data and interpret the results of the methods (JAMK, Curriculum of Data Analytics course, 2022).

Course Contents

- Python data analytics libraries: NumPy, Pandas and Matplotlib
- Data visualization: creation and analysis of descriptors
- Processing of missing values
- Outliers
- Terminology: Average, standard deviation, correlation coefficient and their interpretation
- The concept of probability distribution (particularly normal deviation), confidence interval and hypothesis testing.
- Linear and logistic regression

Machine Learning

Learning Outcomes of Machine Learning Course

The student understands the significance of machine learning in digitalizing operational environment. The student knows the most common machine learning methods, is able to apply them in practice to existing data and interpret the results of the methods (JAMK, Curriculum of Machine Learning course, 2022)

Course Contents

- Mathematical basics in Artificial Intelligence based on Data Analytics
- Most common regression and classification models of guided machine learning and their application in Python programming environment using NumPy, Pandas, Scikit-learn and Keras libraries.
- Processing of Pandas DataFrame object
- Linear regression model
- Logistic regression model
- NLP neural networks
- Accuracy estimation of model
- Support Vector Machine
- K-nearest neighbours algorithm
- K-means clustering
- Time series and recursive neural networks (LSTM)
- Image classification and convolution neural networks

Deep Learning

Learning Outcomes of Deep Learning Course

The student understands the significance of deep learning in the digitalizing operational environment. The student knows about the most common methods of deep learning and how to apply them in practice to existing data as well as interpret the results of the methods (JAMK, Curriculum of Deep Learning course, 2022).

Course Contents

- Various neural networks and their purposes of use
- Work with TensorFlow

- Image recognition
- Prediction
- Reinforcement learning

Project Work on Data Analytics
Learning Outcomes of Project Work on Data Analytics Course

The student understands and masters the various phases of Data Analytics project. The student is able to select the applicable methods for the problem to be solved and apply them to the problem to be solved. The student is able to interpret the obtained results and draw conclusions based on them (JAMK, Curriculum of Project Work on Data Analytics course, 2022).

Course Contents

- Analysis of elective data in Python programming environment
- Includes all stages of data analysis:
- Data preprocessing
- Data description and descriptors
- Selection of a suitable predictive model and its implementation (at least two alternative models)
- Assessment of the accuracy of the predictive models
- Detailed reporting (GitLab)

Project Work on Artificial Intelligence
Learning Outcomes of Project Work on Artificial Intelligence Course

The student understands and masters the various phases of Artificial Intelligence project. The student is able to select the applicable methods for the problem to be solved and apply them to the problem to be solved. The student is able to interpret the obtained results and draw conclusions based on them (JAMK, Curriculum of Project Work on Artificial Intelligence course, 2022).

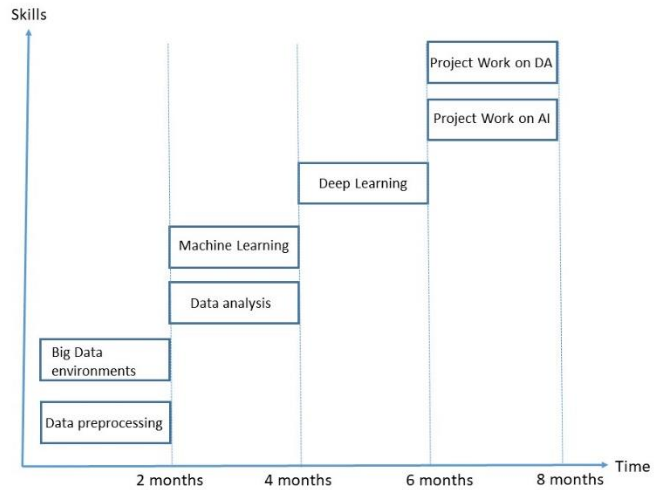
Course Contents

- AI assignment based on optional (elective) data
- Data preprocessing
- Choice of neural network
- Building a neural network
- Choice of training set and test set
- Assessment of results (overfitting/underfitting)
- Detailed reporting (GitLab)

Course Schedule

There are dependencies between the developed courses due to the skills taught in the previous courses. Therefore, the courses must be completed in a certain order and based on a certain time schedule. The individual courses are scheduled to last about 2 months, so the study is relatively intensive. As can be seen from the contents of the courses, the aim has been to teach basic skills throughout the data-driven project. The order of the courses and time schedule is presented in Figure 6.

FIGURE 6
COURSE TIMETABLE AND SCHEDULE



Curriculum Implementations

The curriculum was offered for the first time during August 2019. For the first implementation, only 35 participants were accepted, and the course was full within three minutes and over 100 students were left in a queue to wait for a possible cancellation place. Due to the high demand, the second implementation started with 50 students in November 2019 without any additional enrolment from the queue. The third implementation started in January 2020, and the course was full within eight hours, there were 100 study places and the first 45 study places were reserved in about 2 minutes. The fourth enrolment with 50 participants started in the spring 2020, and more than 100 students were left in a queue to wait for a possible cancellation place.

The next generation of curriculum is under development, and it will be started in 2022 with a few modifications to the courses. The future modifications will be based on the statistics of the courses as well as best practices in data-driven based research projects at Jamk.

RESULTS: THE MAPPING OF PROCESS MODELS, JOBS, AND CURRICULUM COURSES

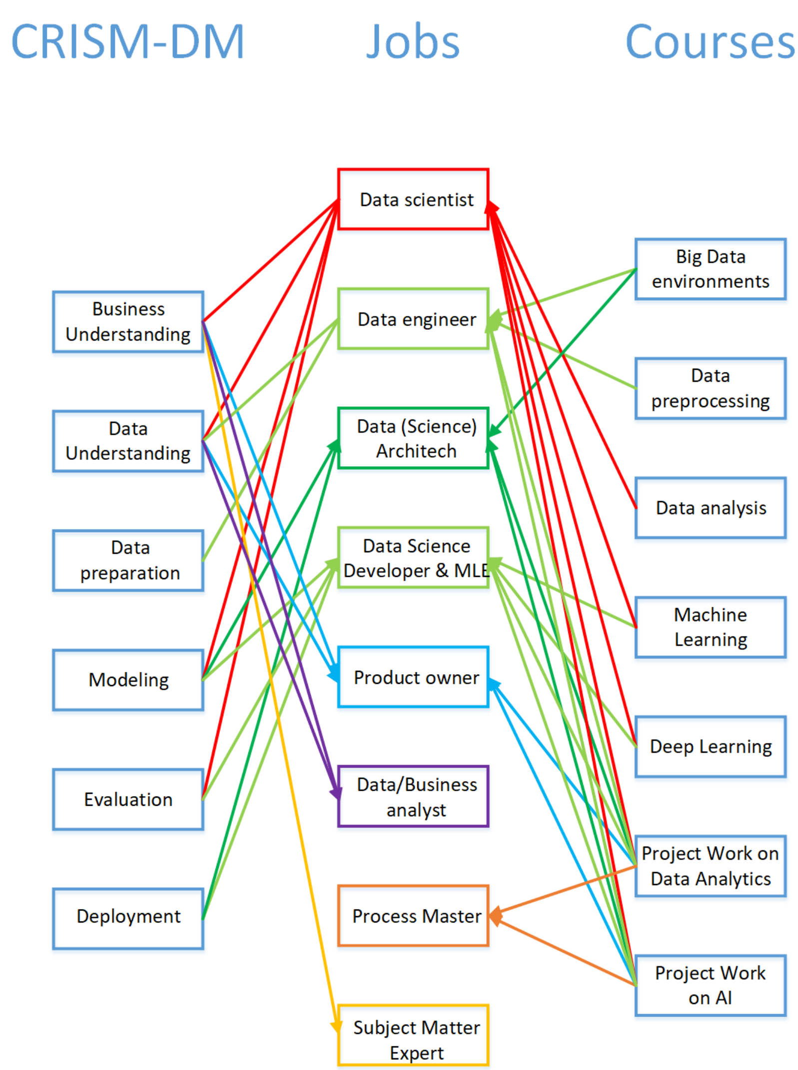
This chapter presents the results, how the phases of most commonly used data-driven project models (CRISP-DM) reflect to the data-driven real-life jobs and their descriptions and required skills, and further how the jobs and the developed courses map together. As mentioned earlier, the courses and their contents are derived from the field studies from different domains and data. With our experience, a wide variety of challenges has been solved and refined into requirements of hands-on skills. Also, the project models used in Jamk's data-driven R&D projects and the good practices derived from them have been included in the studied courses. Correspondingly, the contents of the courses have been mapped to new real-work jobs descriptions, and the listed skills in them can be evaluated due based on the coverage of course competences.

Figure 7 shows the mapping between job titles, CRISM-DM work phases and the curriculum courses. On the left side of Figure 7 the phases of the DRISP-DM process model are shown with how they reflect to the most commonly used real-life job titles based on the required skills. As can be seen in Figure 7, a single job description has an interface to multiple process phases. Correspondingly, the combination of job descriptions and the contents and competencies of the developed courses can be found on the right side of Figure 7.

The curriculum courses start from the initial stages of data-driven projects, starting with the big data environments and how data can be collected from the most commonly used systems or APIs. Next, the data has to be pre-processing, where the data is converted into a format that can be used by either DA, ML, or DL algorithms. These data-driven technologies require in-depth knowledge of algorithms and their application possibilities. Once the data processing and application of different models have been learned, the purpose of the project courses is to gain an understanding of data-driven project management and steering.

The contents of the developed courses consist largely of hands-on data preparation and data-driven algorithm development, the purpose of which is to teach students to practice making DA and AI applications and to produce new skills for IT companies and the challenges of working life.

FIGURE 7
MAPPING OF PROCESS MODELS, JOBS, AND CURRICULUM COURSES



CONCLUSION

In the spring of 2017 Jamk started with the help of internal funding to develop the expertise in DA and AI. Precise knowledge of course contents has been gained hands-on by doing, because every application of

DA and AI differs in some way. Furthermore, the process of data-driven projects varies significantly from a traditional ICT project as well as the use of DA or AI software varies from the use of traditional use of program libraries. During the last years, Jamk has executed data-driven projects and the staff and students have familiarized themselves with different data-driven project models. The experience of data-driven project execution has also been considered. The application of DA and AI requires a profound knowledge of the applicability of various ML and DL methods and neural networks to the problem that is to be solved. Knowledge of theory alone is not enough; DA and AI must be tested in practice. The real-world working life jobs and titles have changed over the last years. The requirements of data-driven jobs set the need for new skills and competences. The lack of good and well-organized university level education material and knowledge of the previously mentioned technology were the major challenges during the development process. Basic and very high-level theoretical courses from DA and AI can be found easily; however, hands-on type implementation courses were missing. Also, the combination of courses which comprise the whole data-driven process is not available. This continuous and iterative process of curriculum development has now been refined and is offered as 30-ECTS education in DA and AI, which started in the fall of 2019 in the Open Studies syllabus at the JAMK University of Applied Sciences.

ACKNOWLEDGMENTS

This work was supported by the Regional Council of Central Finland/Council of Tampere Region and European Regional Development Fund as part of the Data for Utilisation - Leveraging digitalisation through modern artificial intelligence solutions and cybersecurity project of JAMK University of Applied Sciences Institute of Information Technology. The authors would like to thank Ms. Tuula Kotikoski for proofreading the manuscript.

REFERENCES

Alpaydin, E. (2021). *Machine Learning.* MIT Press essential knowledge series.

Brechner, E. (2015). *Agile Project Management with Kanban.* Microsoft Press.

Conway, D. (2010, September 30). *The Data Science Venn Diagram*. Retrieved March 2013, from http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Conway, D. (2010, September 30). *The Data Science Venn Diagram*. Retrieved April 5, 2021, from http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Dickens, D.R. (1977). *Hermeneutics and Ethnomethodology*. Southwestern Sociological Association.

EMC. (2015). *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data.* Wiley Publishing.

Feyyad, U.M. (1996). Data mining and knowledge discovery: making sense out of data. *IEEE Expert*, pp. 20–25.

Foroughi, F.A. (2018). Data Science Methodology for Cybersecurity Projects. *5th International Conference on Artificial Intelligence and Applications.*

Goodfellow, I.B. (2016). *Deep Learning.* MIT Press.

Grady, J.S. (2017). The ambiguity of data science team roles and the need for a data science workforce framework. *2017 IEEE International Conference on Big Data (Big Data).*

JAMK. (n.d). *Curriculum of Big Data Environments course*. Retrieved April 5, 2022, from https://opetussuunnitelmat.peppi.jamk.fi/en/ZJA2020KTDA3/course_unit\\/TTOW14000

JAMK. (n.d.). *Curriculum of Data Analytics course*. Retrieved April 5, 2022, from https://opetussuunnitelmat.peppi.jamk.fi/en/ZJA2020KTDA3/course_unit\\/TTIW0300

JAMK. (n.d.). *Curriculum of Data Pre-processing course*. Retrieved April 5, 2022, from https://opetussuunnitelmat.peppi.jamk.fi/en/TTV2020SS/course_unit/TTC8030

JAMK. (n.d.). *Curriculum of Deep Learning course*. Retrieved April 5, 2022, from https://opetussuunnitelmat.peppi.jamk.fi/en/ZJA2020KTDA3/course_unit\\/TTOW14000

JAMK. (n.d.). *Curriculum of Machine Learning course*. Retrieved April 5, 2022, from https://opetussuunnitelmat.peppi.jamk.fi/en/TTV2021SS/course_unit\\/TTC8050
JAMK. (n.d.). *Curriculum of Project Work on Artificial Intelligence course*. Retrieved April 5, 2022, from https://opetussuunnitelmat.peppi.jamk.fi/en/ZJA2020KTDA3/course_unit\\/TTOW1500
JAMK. (n.d.). *Curriculum of Project Work on Data Analytics course*. Retrieved April 5, 2022, from https://opetussuunnitelmat.peppi.jamk.fi/en/ZJA2020KTDA3/course_unit\\/TTOW1600
Kelleher, J. (2019). *Deep Learning.* MIT Press.
Lee, W.-M. (2019). *Introduction to Machine Learning.* John Wiley & Sons, Ltd.
Microsoft Docs. (n.d.). *What is the Team Data Science Process? - Azure Architecture Center*. Retrieved April 5, 2022, from https://docs.microsoft.com/en-us/azure/architecture/data-science-process/overview
Pichler, R. (2010). *Agile Product Management with Scrum: Creating Products that Customers Love.* Addison-Wesley Professional.
Pries, K.H. (2010). *Scrum Project Management.* USA: CRC Press, Inc.
Rantonen, M.A. (2020). Conversion of Emerging ICT-Technology into Curriculum Courses. *Journal of Strategic Innovation and Sustainability*, pp. 70–77.
Rollins, J.B. (2015, June). Foundation Methodology for Data Science. *IBM Analytics*. Retrieved April 5, 2022, from https://tdwi.org/~/media/64511A895D86457E964174EDC5C4C7B1.PDF
Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, pp. 13–22.
Silver, A. (2018, September 27). *The Essential Data Science Venn Diagram*. Towards Data Science. Retrieved April 5, 2022 from https://towardsdatascience.com/the-essential-data-science-venn-diagram-35800c3bef40
TensorFlow. (n.d.). *An end-to-end open source machine learning platform*. Retrieved April 5, 2022, from https://www.tensorflow.org/
van der Aalst, W.M. (2016). *Process Mining: Data Science in Action.* Heidelberg: Springer.
Whitaker, D.E. (2019). *Point of View: Data science for the scientific life cycle*. eLife Sciences Publications.