Conversion of Emerging ICT-Technology into Curriculum Courses

M. Rantonen
JAMK University of Applied Sciences

M. Karjalainen
JAMK University of Applied Sciences

The aim of this paper is to describe the process of how to develop course content for a new challenging and emerging ICT technology in a higher education organization. The structure of the developed data analytics (DA) and Artificial Intelligence curriculum (AI) as well as the contents of the courses will be presented precisely. In spring 2017, the Institute of Information Technology at JAMK University of Applied Sciences started to develop its competence in DA and AI. 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.

Keywords: Artificial intelligence, Data analytics, curriculum courses, Big Data, emerging technology, virtual courses

INTRODUCTION

AI has appeared in the field of ICT as an emerging technology in recent years, and the term seems to be on almost everyone's lips. High expectations have been set for AI, which has reached a level of maturity where it can be applied to almost all industry areas.

JAMK University of Applied Sciences (JAMK) started in the spring of 2017 with the help of internal funding to develop the expertise in data analytics (DA) and Artificial Intelligence (AI). The Institute of Information Technology at JAMK has invested in developing and applying its expertise in DA and AI for different areas and substances. The emphasis has been placed on the practical application of open source DA and AI products, either based on available open data or on real business-based problems. The challenge of bringing in the students to practice hands on work in the above-mentioned fields has been caused by the lack of educational courses at JAMK. The research team's approach to learning things has been hands on training, which has proven to be a very good solution. Knowledge has been gained precisely by doing, because every object of both application areas differs in some way from another object. The application of DA and AI differs significantly from the traditional use of program libraries in that the application of AI requires a profound knowledge of the applicability of various machine learning (ML) methods and neural networks to the problem to be solved. Knowledge of theory alone is not enough; developed AI methods must be tested in practice. In addition to the practical hands on work of

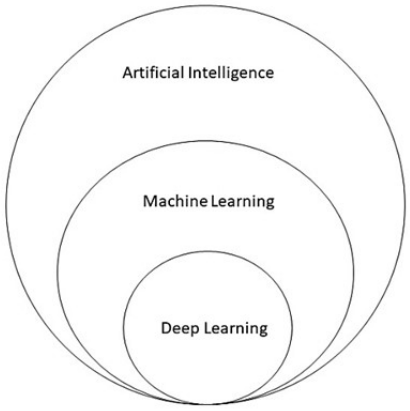
the research team, it was noticed that the prolific knowledge can be transformed into courses that can be used in the future for university-level teaching.

METHODOLOGY

Modern information technology, incremental computing power, and online digitalization have opened new opportunities to utilize automatically collected and stored data from various sources. The exponentially growing amount of data is referred to as a Big Data and the traditional software and tools are not anymore applicable to process enormous amounts of unstructured data. Furthermore, the traditional computational pattern discovery of data mining is replaced with modern data analytics, ML and deep learning (DL) methods.

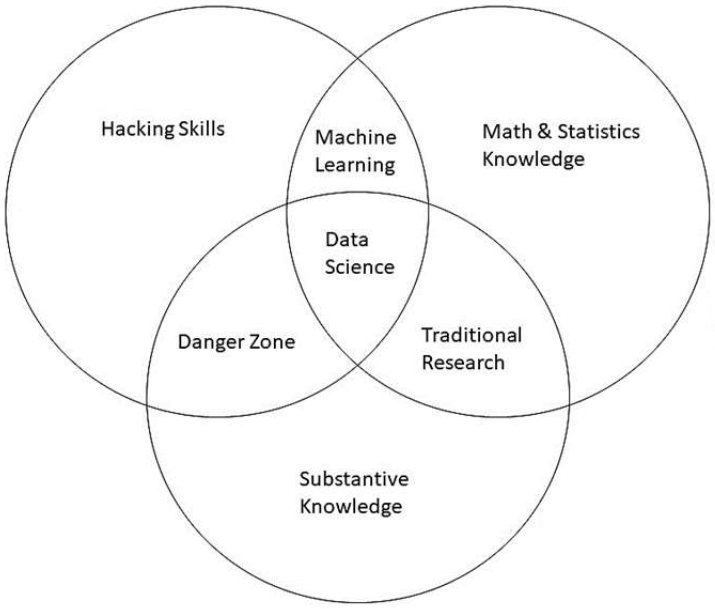
AI has existed over many decades, and the field is enormously wide (Goodfellow, 2016). AI can be viewed as an umbrella term that contains ML and DL. The ML is a subset of AI; meanwhile, DL, in turn, is a subset of ML (Fig. 1). The term DL refers to artificial neural networks (ANN) with complex multilayers (Abiodun, 2018).

FIGURE 1
A VENN DIAGRAM SHOWING THE RELATIONSHIP BETWEEN AI TERMINOLOGY



Respectively, data science/analytics is a relatively new industry, albeit the fact that its components have been around for a long time. Venn diagram (Fig. 2) is used to understand that data science is a combination of several disciplines. In this Venn diagram, the three components are hacking skills, math & statistics knowledge, and substantive expertise (Conway, 2013), (Silver, 2018).

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



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 lots of different kind of a very extensive skill set. JAMK's approach to learning things has been hands on training, which has proven to be a very good solution. Knowledge has been gained precisely by developing different kind 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 is not enough; 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.

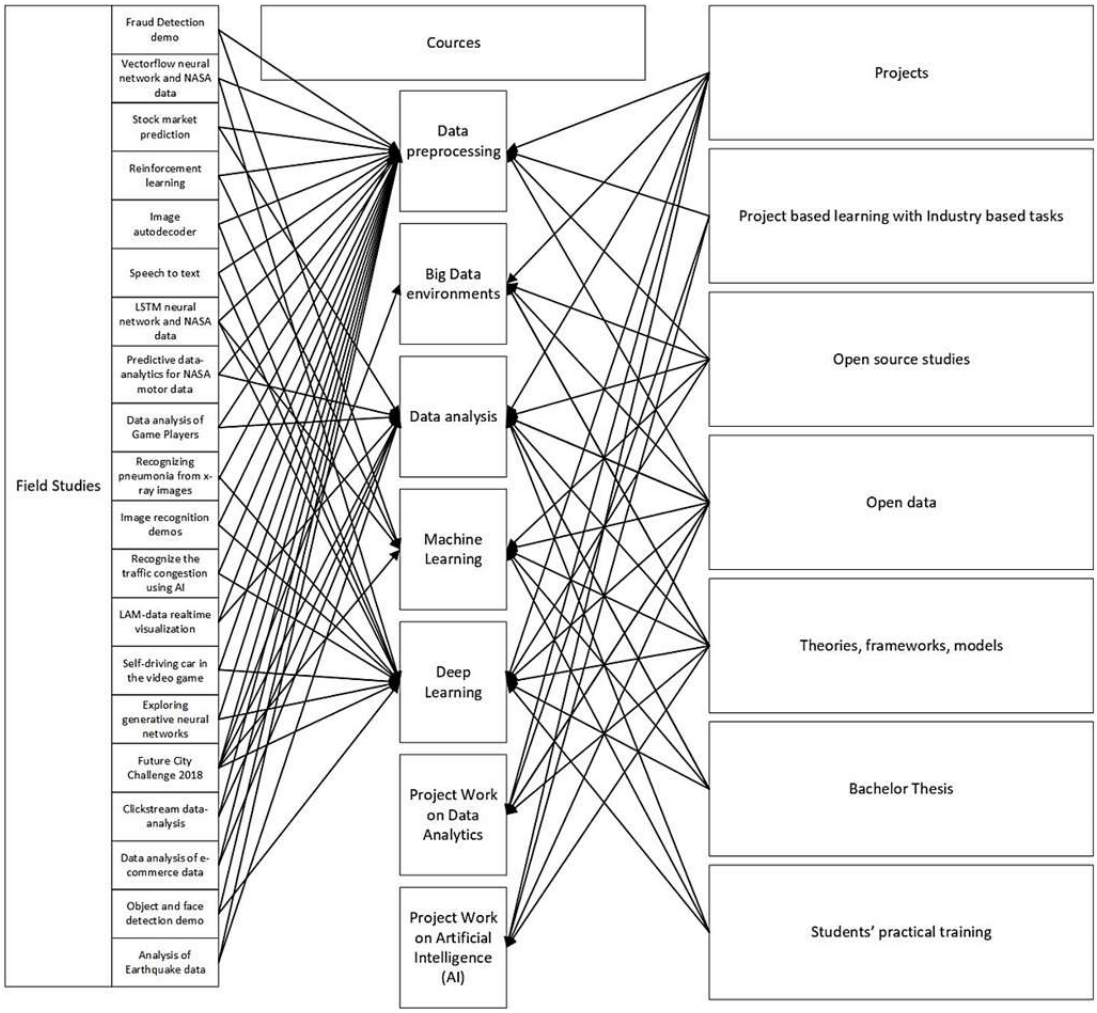
RESULTS

Development Process and Framework of Education Courses in DA and AI

JAMK has responded to the lack of practically oriented educational courses in the area of DA and AI by introducing a 30-ECTS course offered by the Open University of Applied Sciences in autumn 2019. At the curriculum course developed there has been used the hermeneutic circular method (Dickens, 1977). In this case, by the literature review and general study from the area have been generated a general understanding of the area applicable. After review phase the understanding have deepened through a series of practical tests that have applied prior understanding of the area. In this way the curriculum frame has been built through the in-depth knowledge through exercises. Furthermore, Cross-industry standard process for data mining, known as CRISP-DM (Shearer, 2000) is an open standard process model that describes common approaches used in data mining, data analytics and deep learning projects. CRISP-CM has been used the map special skills and understanding using the developed framework to the specific curriculum course.

The content of the courses is designed based on the experience gained by a research team in the past 2 years regarding DA and AI. The course table consists of seven different completely virtual courses. All courses contain a lot of practical exercises, as is customary in Bachelor of Engineering type of education. Figure 3. illustrates the Framework of course development, describing the operating environment, the different sources and the working methods with which the courses have been developed. Based on the field's studies and different dataset, the Framework is applied and courses are defined to cover all needed skills.

FIGURE 3
FRAMEWORK OF COURSE DEVELOPMENT



Field Studies

JAMK's Data Analysis and Artificial Intelligence website (Artificial intelligence and data analysis, 2020) has published an example application with source code and data sets for those interested in it. The purpose of the sites is to provide information on applications, capabilities and examples of DA and AI that anyone can test according to their own interests. The idea behind the application examples was to use different kind of data and DA and AI methods as well as to find the core competencies of the development of DA and AI applications.

At the field study of fraud detection demo we took the credit card fraud detection as a case by using data classification based on neural networks. The focus of this field study was to recognize misuses of

credit card using the information of transactions where the timestamp and amount of money are not anonymized (Credit card fraud demo, 2018). At vectorflow neural network and NASA data field study case, we tested vectorflow neural network-based prediction of engine failure in NASA motor data (Vectorflow RUL prediction demo, 2018). Furthermore, LSTM (Long short-term memory) neural network and Predictive data-analysis are used to predict the engine failure base on NASA motor data (Reamaining Useful Life (RUL) prediction, 2018) and (NASA motor data, 2018). In Predictive data analysis case, a linear regression and perceptons are used to predict the engine failure. At stock market prediction field study, we tested stock market prediction by using LSTM neural network to determine when to sell, buy or hold the specific stock based on the next day's closing price (Stock market prediction, 2018). At reinforcement learning we took the reinforcement learning method demo to teach worm movements in Worm game (Reinforcement learning, 2018). At image autodecoder use case, the neural network has been applied as an encoder to compress an image to floating point values. The decoder attempts to decompress their values back to an image resembling the original image (Autoencoder example, 2018). At Speech to text field experiment, speech to text translation where the learning scripts utilize the Tensorflow Dataset and Estimator APIs (Speech to Text with Neural Networks, 2018). Data analysis of Game Players study, data analysis of Battlegrounds players based on the movements and activities of players during the whole game duration were analyzed. Clustering method was used to categorize players' activity and success (Player Unknown´s Battlegrounds data analysis, 2018). At case recognizing pneumonia from x-ray images we used convolutional neural network (CNN) and transferred learning to recognize pneumonia with a small training data set (Predicting pneumonia from chest x-ray images with convolutional neural network, 2019). At case Image recognition demo, we use easy-to-use image recognition demos that work in a web browser. The open sets were used to teach the neural network developed using Keras (library). The loading of the model and the prediction were established by Tensorflow.js (Image recognition demos, 2020). At field study Recognize the traffic congestion using AI, AI based traffic congestion recognition was used. The road weather camera images of Finnish Transport Agency were fetched and the number of objects (cars etc.) was calculated using the neural network. The number of objects is visualized as a heat map on Google Maps (Traffic congestion demo, 2018). At field study LAM-data real time visualization, visualization demo application showed how to use deck.gl and react-map-gl to visualize LAM traffic data from Finnish Transport Agency's open data API (LAM-data visualization, 2019). At field study Self-driving car in the video game, the neural network is used to teach the self-driving car to drive independently on the video game tracks (Computer driven car in a simulated environment, 2018). At field study Exploring generative neural networks (GAN) we studied how neural networks can be used for creative purposes, such as synthesizing images from scratch. We explored generative models as a method of enhancing and drawing realistic human faces (Adversarial Generative models, 2018). At field study Future City Challenge 2018 JAMK, a student team made a proposal to the Future City Challenge (FCC) in Jyväskylä. The aim was to embrace resource wisdom city (Future City Challenge, 2018). At field study Clickstream data-analysis, clickstreams, also known as click paths, we study how the route visitors choose when clicking or navigating through a web site. Using data analytics, we examined the online shopping clickstream associated with the purchasing of products and sites from which Wikipedia has been accessed (Click path (Clickstream), 2018). At field study Object and face detection demo, object recognition used neural network either from the still images or video stream. Testing face recognition was integrated into the object recognition if the object is recognized as a person. Furthermore, the object recognition was integrated into the Nao robot (Face detection with TensorFlow object detection API, 2018). At field study Analysis of Earthquake data, data analysis and 3D visualization of earthquake data were performed (Clustering earthquake data, 2018).

Research and Development and Co-operation with Students

At JAMK, R&D in DA and AI in project form has started to grow as the expertise increases. Project preparation and application have been made possible by in-depth knowledge of DA and AI, both in theory and in practice, which enables us to make good applications to different funders. R&D at JAMK can be called applied R&D, which differs significantly from academic research in the university world. Applied

R&D uses fully open source products and tools, which means applying existing ready methods to business problems. There is more academic and basic research in university research, going much deeper in algorithms or neural network architecture as in the applied R&D. 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 business problem. The application of AI has been facilitated by the introduction of AI libraries by various commercial operators, perhaps best known as Google's Tensorflow (Tensorflow, 2020). In DA, the Python programming language is selected due to its well-known capability in conducting scientific calculation.

A research team consisting of staff and students is responsible for developing JAMK's expertise in DA and AI. Almost 20 Bachelor of Information and Communication Technology students have participated in the research group by completing a training or thesis in the field of DA and AI. Currently, 7 of them are working at JAMK as project staff, either in artificial intelligence or data analytics projects.

Projects and Project Based Learning

In the spring of 2018, projects funded by the European Regional Development Fund (ERDF) were launched in the field of New Knowledge in Data Analysis and Business and Investments in a Data Security Development Environment. The projects aim to add value and new innovations to enterprise dark data through data analytics and machine learning methods. With the projects, the research team has built a secure data analytics and machine learning development environment. The project participants form a cluster of companies in the energy sector or closely related industries. The companies involved in the project are Fingrid, Alva (earlier Jyväskylän Energia), Landis+Gyr and C2Smartlight. The location of companies in the same field of industry enables data analytics and its utilization and development throughout the value chain: Production <-> Distribution <-> Consumption <-> Customers. The projects also include cases of common interest between companies seeking to combine corporate data and find synergies. In addition, JAMK has also carried out a research project directly funded by the National Defense Science Advisory Board (MATINE): Using Artificial Intelligence to Detect Anomalies in Web Traffic. In addition to these projects, JAMK uses a lot of project-based learning in teaching, where we have collaborated with several companies and gained experience of the need for different environments for artificial intelligence and data analytics.

Open Source Tools and Data

JAMK has deliberately focused on using open source solutions, the release of which has accelerated the development of artificial intelligence in recent years. In addition, these have allowed for the full allocation of funding to skills development, without burdening licensing and software fees. Essential to vendor independence is also the avoidance of the so-called vendor lock, *i.e.* commitment to one and only cloud service or software provider.

The development of AI requires data to train ML models or neural networks. The amount of data can be substantial depending on the application. The most used environments are cloud-based, where scalable computing time, storage, and memory can be purchased. While the consumption of cloud computing, storage and memory is compared and pricing may be quite clear, it is very difficult to predict the overall cost in DA or AI projects. For example, when training a neural network, the cost of computing capacity may come as a surprise. Another issue that has emerged in business collaboration is data security, security and legal issues for cloud-based DA and AI environments. This has been the most common issue when it comes to the disclosure of corporate data. Companies want a guarantee that data is stored properly, and they may not want to put it in the cloud. An alternative is to build your own DA and AI environment. Costs of your own computing environment may be high; however, after that the costs consist mainly of maintenance and power consumption.

The clear challenge for practical implementation has been data, or indeed the lack of it. Both DA and AI need to have enough data and preferably still, enough quality data. Without data you cannot optimize DA algorithms or teach ML algorithms. For many companies, data sharing poses a challenge for three main reasons:

1. Data informs about their critical business and the company does not dare to divulge their data despite confidentiality agreements.
2. Data can be stored in many locations, meaning it is in multiple locations and very old systems, which it may be very cumbersome to collect it.
3. Data quality may be poor; i.e. there may be many missing values, duplicate values or values that do not exceed the rating scale. With such poor-quality data, it is almost impossible to make a good ML model

DA and AI Curriculum Courses

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 diversity of the use cases provides a strong insight how the curriculum must be structured, and what kind of knowledge has to be included (Fig. 3).

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 preprocessing, 3 credits
- Big Data environments, 5 credits
- Data analysis, 4 credits
- Machine Learning, 5 credits
- Deep Learning, 5 credits
- Project Work on Data Analytics, 4 credits
- Project Work on Artificial Intelligence (AI), 4 credits

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

CONCLUSIONS

Precise knowledge of course contents has been gained hand on by doing, because every application of DA and AI differs in some way. Furthermore, the process of DA and AI 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. The application of DA and AI requires a profound knowledge of the applicability of various ML 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.

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 at the Open University of Applied Sciences. In the first implementation, only 35 participants were accepted, and the course was full within three minutes. The second enrolment started on 18 November 2019 with 100 study places and the course was full within eight hours. The first 45 study places were reserved in about 2 minutes. In addition, JAMK will launch a master's degree program in AI and DA with tuition in English in the autumn of 2020.

In the field of technology, the rapid development of technologies presents challenges for keeping teaching up to date. Basic research in the field of artificial intelligence has a long tradition, and we are now at the stage where technology maturity allows for a large-scale application. Thus, artificial intelligence and data analytics must be integrated into a modern engineering education. Educationally, taking over the aforementioned technology is a major challenge. By following the model presented in this paper, the know-how generated by practice can be generated into course structures. The use of the model is of prime importance, especially in the field of applied engineering, where the course content must provide an opportunity for the student to learn with hands on exercises and training.

REFERENCES

Abiodun, O. I. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11).

Adversarial Generative models. (2018, June). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/generative-adversarial-model-experiments

Artificial intelligence and data analysis. (2020, March). Retrieved from https://jamk.fi/en/ai

Autoencoder example. (2018, February). Retrieved from ttps://gitlab.labranet.jamk.fi/data-analysis-and-ai/autoencoder-demo

Click path (Clickstream). (2018, June). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/da-documentation/tree/master/clickstream

Clustering earthquake data. (2018, May). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/earthquake

Computer driven car in a simulated environment. (2018, June). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/tmnf-car

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

Credit card fraud demo. (2018, February). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/credit-card-fraud-demo

Dickens, D. (1977). *Hermeneutics and Ethnomethodology.*

Face detection with TensorFlow object detection API. (2018, May). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/object-detection-demo

Future City Challenge. (2018, April). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/FCC

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

Image recognition demos. (2020, March). Retrieved from https://aida2.labranet.jamk.fi/

LAM-data visualization. (2019, October). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/lam-station-visualization

NASA motor data. (2018, March). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/nasan_moottoridata

Player Unknown´s Battlegrounds data analysis. (2018, May). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/PUBG-data-analysis

Predicting pneumonia from chest x-ray images with convolutional neural network. (2019, May). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/pneumonia-cnn

Reamaining Useful Life (RUL) prediction. (2018, February). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/nasa-rul-prediction-lstm

Reinforcement learning. (2018, February). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/reinforcement-learning-demo

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, pp. 13-22.

Silver, A. (2018, September). *The Essential Data Science Venn Diagram*. Retrieved from https://towardsdatascience.com/the-essential-data-science-venn-diagram-35800c3bef40

Speech to Text with Neural Networks. (2018, October). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/speech-to-text

Stock market prediction. (2018, February). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/stock-market-prediction

Tensorflow. (2020, March 17). *An end-to-end open source machine learning platform*. Retrieved from https://www.tensorflow.org/

Traffic congestion demo. (2018, June). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/traffic-congestion

Vectorflow RUL prediction demo. (2018, February). Retrieved from https://gitlab.labranet.jamk.fi/data-analysis-and-ai/nasa-rul-prediction-vectorflow