Teaching Practical Machine Learning Concepts to Professionals and Students: An Integrated and Interdisciplinary Qualification Project

Tobias Hagen
Business and Industrial Engineering
Offenburg University of Applied Sciences
Offenburg, Germany
tobias.hagen@hs-offenburg.de

Tobias Lauer
Electrical Engineering, Medical
Engineering and Computer Science
Offenburg University of Applied Sciences
Offenburg, Germany
tobias.lauer@hs-offenburg.de

Volker Sänger
Media and Information
Offenburg University of Applied Sciences
Offenburg, Germany
volker.saenger@hs-offenburg.de

Klaus Dorer
Electrical Engineering, Medical
Engineering and Computer Science
Offenburg University of Applied Sciences
Offenburg, Germany
klaus.dorer@hs-offenburg.de

Stephan Trahasch
Electrical Engineering, Medical
Engineering and Computer Science
Offenburg University of Applied Sciences
Offenburg, Germany
stephan.trahasch@hs-offenburg.de

Abstract—Machine learning (ML) has become highly relevant in applications across all industries, and specialists in the field are sought urgently. As it is a highly interdisciplinary field, requiring knowledge in computer science, statistics and the relevant application domain, experts are hard to find. Large corporations can sweep the job market by offering high salaries, which makes the situation for small and medium enterprises (SME) even worse, as they usually lack the capacities both for attracting specialists and for qualifying their own personnel. In order to meet the enormous demand in ML specialists, universities now teach ML in specifically designed degree programs as well as within established programs in science and engineering. While the teaching almost always uses practical examples, these are somewhat artificial or outdated, as real data from real companies is usually not available. The approach reported in this contribution aims to tackle the above challenges in an integrated course, combining three independent aspects: first, teaching key ML concepts to graduate students from a variety of existing degree programs; second, qualifying working professionals from SME for ML; and third, applying ML to real-world problems faced by those SME. The course was carried out in two trial periods within a government-funded project at a university of applied sciences in south-west Germany. The region is dominated by SME many of which are world leaders in their industries. Participants were students from different graduate programs as well as working professionals from several SME based in the region. The first phase of the course (one semester) consists of the fundamental concepts of ML, such as exploratory data analysis, regression, classification, clustering, and deep learning. In this phase, student participants and working professionals were taught in separate tracks. Students attended regular classes and lab sessions (but were also given access to e-learning materials), whereas the professionals learned exclusively in a flipped classroom scenario: they were given access to e-learning units (video lectures and accompanying quizzes) for preparation, while face-to-face sessions were dominated by lab experiments applying the concepts. Prior to the start of the second phase, participating companies were invited to submit real-world problems that they wanted to solve with the help of ML. The second phase consisted of practical ML projects, each tackling one of the problems and worked on by a mixed team of both students and professionals for the period of one semester. The teams were self-organized in the ways they preferred to work (e.g. remote vs. face-to-face collaboration), but also coached by one of the teaching staff. In several plenary meetings, the teams reported on their status as well as challenges and solutions. In both periods, the course was monitored and extensive surveys were carried out. We report on the findings as well as the lessons learned. For instance, while the program was very well-received, professional participants wished for more detailed coverage of theoretical concepts. A challenge faced by several teams during the second phase was a dropout of student members due to upcoming exams in other subjects.

Keywords—Machine learning, qualification, interdisciplinary

I. INTRODUCTION

A. Machine Learning is Interdisciplinary

Machine Learning (ML) is an interdisciplinary field that requires knowledge and abilities from different disciplines. ML knowledge includes mathematics – in particular statistics – to understand and perhaps to adjust the algorithms used. In order to apply and to use these algorithms, programming abilities and skills in the area of computer science, such as databases or software engineering, are necessary. These are more or less the technical basics. However, in order to know where and how to apply ML, and to foresee the benefits of ML, practical working experience in the application field is absolutely inevitable: problems and shortcomings of application areas such as transportation, e-commerce or production planning must be well understood. But how can one acquire these abilities together? Perhaps computer scientists have studied some mathematics and statistics during their education, but they do not know too much about the various application fields of ML. And normally experts working in production planning, transportation or e-commerce (often managers or engineers) are far away from mathematics and computer science. But they are the ones to identify new application areas for ML.

B. Current State of Machine Learning Curricula

At the authors' home institution, a university of applied sciences in south-west Germany, several ML modules exist in computer science master's programs. One or the other module can also be found in the fields of media and of business administration at the master's level. In the near future, there will be a bachelor degree program in "Applied Artificial Intelligence", where ML will play a central role. Hence, up to now, there has been no interdisciplinary teaching concept.

Worldwide, there have been many attempts to integrate ML in teaching, meanwhile not only in master's programs, but more and more in undergraduate programs [3, 5], sometimes even in high school [2] or for children [4]. They all act in the basic tenor that ML needs to be taught based on example usage scenarios. A study on educational content of ML teaching confirms this [1]. A major problem for universities with such application scenarios is that these scenarios are always somewhat artificial, as they cannot really relate to a current scenario from a company: no company would release the corresponding real-world data.

Taken together, it seems like ML teaching could strongly benefit from an interdisciplinary approach.

C. Industry Demand

It is obvious that ML has become highly relevant in recent years. ML specialists are sought after urgently. For example, the study reported in [6] finds that, in Germany alone, 85,000 university graduates with skills in data analysis are missing, especially in the area of engineering and medicine. Another 10,000 IT specialists in the field of data analysis are required. The breadth of ML training is seen as a crucial factor, because often both engineering and ML skills are required – again, interdisciplinarity is essential.

According to the above study, SMEs in particular need to catch up with ML techniques. SMEs often show interest as soon as they see specific applications with entrepreneurial added value. Support measures should therefore allow employees to use ML in general, for example by providing more information on the use and benefits of ML-based techniques.

Especially in SMEs, professionals like engineers are very important because they drive the companies' products and hence they do not have much free time. On-the-job further education is possible, but as time is short and their focus is on engineering in their application area, they will never advance ML projects.

Another problem with the use of ML is the availability of data: companies do not release their data for ML projects because they fear disadvantages in case the data becomes known to the public or competitors.

D. Research questions

As we can see, machine learning is a multidisciplinary field of expertise that is considered highly relevant to a wide range of practical applications. Unfortunately, corporate employees generally do not have the knowledge to use machine learning in their respective disciplines. Computer scientists and mathematicians in turn do not have the necessary application domain knowledge.

The work reported here addresses this problem and has been carried out within an interdisciplinary research project funded by the German ministry of Research and Education (BMBF). The central research question is: "How can potential of machine learning be leveraged for medium-sized companies?" Within the framework of the project, the following more detailed questions of the research agenda are examined:

- What type of know-how must be transferred to the companies?

- How are tasks assigned to the different disciplines and how is their cooperation in projects regulated?

- What results can be achieved in these interdisciplinary projects?

- How can initial results be built so that lasting progress can be made and machine learning can be permanently integrated into business processes?

The remainder of this paper is structured as follows. Section II gives a detailed outline of the concept of the qualification program. In Section III, we describe the actual implementation of the project during the academic years 2018 and 2019. Section IV lists the results and lessons learned during the project, before we give some concluding remarks in Section V.

II. CONCEPT

A. Overall goals

The project has the goal to overcome the shortcomings of today's ML industry: Practitioners know their own field but do not know much about ML. At the same time new ML projects should be driven from the application area, hence from practitioners in these application fields, such as transportation, medicine, e-commerce – in fact, every area where large amounts of data or complex data accumulate. Therefore, the goal is to qualify practitioners so that they themselves are able to discover, elaborate and use their potential in their field of application. For the subsequent implementation of machine learning projects within the companies, interdisciplinary teams of machine learning specialists and experts from the respective application domain are responsible.

Therefore, the central aim of the qualification program is to qualify practitioners from companies – mainly SMEs in the region – and master students from different degree programs of our university in the field of ML. The focus lies on the joint qualification of these two target groups in order to enable interdisciplinary work across subject, study and departmental boundaries, and to develop ideas for intelligent data products and services and to implement them with the help of ML. Students have their focus on learning new technologies, while practitioners are specialists in their application area: they know problems in practice and want to learn new solutions and approaches to solve them. Together these groups can apply the new technologies of ML and learn how to find new practical solutions with ML approaches. It is therefore very important to recruit people from different application areas. Hence, entry barriers are kept low in order to motivate persons of different qualification, background and age to join the project.

Students and practitioners have different backgrounds. Students are on campus, they can earn credits for the course (and hence need to take a formal exam) and usually have little knowledge on actual problems in companies, while the situation is vice versa for practitioners. Hence, the qualification program is divided into two phases (see Fig. 1). In phase 1 the two target groups learn in separate tracks about ML – concepts and solutions. This phase is called "Machine Learning *for* Practice". In phase 2 they jointly carry out machine learning projects. This phase is called "Machine Learning *in* Practice".
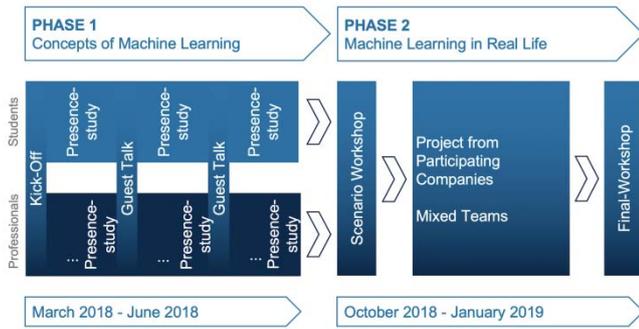
Fig. 1   Competences acquired in the two phases.

For the skills to be acquired during the qualification, we distinguish between technical and interdisciplinary competencies. Technical competencies concern mostly basic knowledge of ML and the general ability to apply this knowledge to simple problems (see also Table I):

- TC1: The graduates have knowledge and understanding of the important basics of ML.

- TC2: They are able to independently acquire new contents in the field of ML as well as to develop their own ideas.

- TC3: They are capable of naming and interpreting machine learning terminologies, boundaries and doctrines.

- TC4: They can independently use machine learning techniques and tools to solve practical tasks.

Interdisciplinary competencies include higher level skills: they are practice-oriented and based on the technical competencies (see also Table I):

- IC1: The graduates are able to apply their problem-solving skills in new application-oriented situations that are in a broader or multidisciplinary context.

- IC2: They can autonomously integrate new knowledge, assess it in a differentiated manner and make well-founded decisions, even in complex contexts, which take into account not only scientific and technical but also social and ethical aspects.

TABLE I.   COMPETENCES ACQUIRED IN THE TWO PHASES

Competence	Phase 1: ML for Practice	Phase 2: ML in Practice
TC1	X	
TC2	X	
TC3	X	X
TC4	X	X
IC1		X
IC2	X	X
IC3		X
IC4		X

- IC3: They have the methodical tools to work constructively and motivated in a team or independently in application-oriented ML projects.

- IC4: They have the ability to communicate their ML solutions in a clear and unambiguous way, to exchange ideas with experts and lay people about ideas, problems and solutions.

B. Phase 1: Machine Learning for Practice

In phase 1, the basics and methods of ML are taught. The goal is that at the end of the first phase practitioners and students have comparable ML competences and a common frame of reference and knowledge, taking into account the heterogeneity of the group of participants for further qualification. Since the prerequisites and the time budget of the two target groups are heterogeneous, the first qualification phase for the practitioners and the students takes place in different qualification paths with target group-specific learning contents that take into account the different requirements. The paths are referred to as student track and employee track.

Students attend regular classes and lab sessions but are also given access to e-learning materials, whereas the working professionals learn exclusively in a flipped classroom scenario: E-learning content and courses on ML with the subject-specific topics and methodological competence form the basis of the qualification. With the help of e-learning content such as video lectures, online quizzes, the learners are able to get acquainted with the basics of ML whenever they have the time, independently of each other – at home or at the company in free time slots, and in their own pace and learning rhythm. E-learning is supplemented by face-to-face lab sessions with practical exercises consisting of data sets and practical questions. It should be noted that the number of face-to-face sessions for practitioners is significantly lower (around 50 %) than that for students – students have more time to learn and are used to it. Employees have less time and the main goal for them is to understand the principles of ML to create ideas of ML applications in their application fields and companies.

Both tracks share guest lectures given by experts from industry and research to present specific relevant projects. Even though these lectures are held during the more theoretically oriented phase 1, they offer a way to form links between the knowledge learned and practical experience and problems and to develop ideas for future practical projects. Additionally, they provide an opportunity to get in touch and exchange ideas with the speakers.

In summary, the focus of the qualification concept in the first phase "Machine Learning for Practice" is on the knowledge transfer of ML concepts and procedures. The didactic concept is based on the principles of blended learning with an inverted classroom lab. At the end of phase 1, participants should have the following skills:

- They have acquired in-depth knowledge of ML and a methodological repertoire of ML techniques.

- They are able to use methods of data preparation and exploratory data analysis using descriptive statistics as well as visualizations

- They know and are able to apply and compare supervised and unsupervised ML procedures. These

include regression, association analysis, classification methods, clustering, as well as neural networks and deep learning.

- They understand the algorithms in such detail that they can perform possible parameterizations and hyper-parameter optimizations

- They know specific quality functions for comparing the procedures and are able to compare and evaluate the procedures on the basis of quality characteristics.

C. Phase 2: Machine Learning in Practice

In the practical phase, the focus is on the combination of theory and practice as well as the linking of the learning places university and enterprise as the later work place for the students. This phase involves the transfer and consolidation of the participants' ML expertise, in which they apply and develop the knowledge acquired in the first phase in a professional and entrepreneurial context. The participants apply their ML know-how situationally and adapt it to the existing conditions in practice and to the company goals. Student and employee tracks are dissolved and new interdisciplinary teams are formed, consisting of students and employees from companies, working together on a problem from a company.

The work done in phase 2 is structured according to the CRISP-DM standard [7], possibly in several iterations. In this phase, participants not only apply the ML concepts from phase 1 in the enterprise environment, but in addition, they experience the typical real-world problems such as availability and quality of data or the applicability of algorithms. Furthermore, the economics of an ML project in the practical environment has to be taken into account. These aspects are essential success factors for any ML project in the real world.

It is very important that all participants can learn from the experience gained across projects, not only from their own project. Hence, from the coordination a clear structure is defined:

(1) Phase 2 starts with a kick-off meeting in which the first steps for each project are planned and discussed.

(2) In several intermediate workshops with all participants, the current state and progress, problems encountered and next steps are presented, analyzed and discussed.

(3) A closing session with a presentation of the final results and the lessons learned rounds up phase 2.

(4) Templates for the presentations to each meeting are predefined and have to be used by every project group, so that all the participants can learn from each project. Hints and experiences from other projects can be exchanged.

III. IMPLEMENTATION

A. Preparation

1) Recruitment of Participants

The project and a call for participation was announced via the university's usual internal and external communication channels: There was an announcement on the university homepage and a mailing addressed to all students and a mailing to existing contacts and networks of the participating professors.

In addition, students in some of the master courses were directly addressed in order to be able to answer questions directly. A specific website with all information about the project was set up and a supplemented with a flyer. Due to limited capacity and in order to allow better planning, a sign up process with two steps was implemented: In a first step, a company interested in participating in the project had to tell how many participants they would like to send (minimum/maximum). Students had to apply formally for participation. After being accepted, all participants had to sign up individually in a second step. Another goal of this procedure was to increase the liability of participation: As there were no course fees (due to public funding of the project), the main intention was to reduce the risk of no-shows and early dropouts, especially after phase 1.

2) Participants from Companies

Due to high interest in the project, especially in the first round, there was a limit of two participants per company, but no interested company was rejected. Table II summarizes the participants from companies for both iterations. A decrease in the overall interest from iteration 1 to iteration 2 has been observed. This may be interpreted as a kind of saturation effect.

TABLE II. PARTICIPANTS FROM COMPANIES IN ITERATIONS 1 AND 2.

Iteration	Companies	Pre-registered	Accepted	Drop-outs	Certified
1	17	42-48[b]	26	1	25
2	11	18	18	2	16
Total	25[a]	60-66[b]	44	3	41

[a] Three companies were part of both iterations
[b] A range with maximum and minimum number of participants per company was possible

Most participants came from the area of computer science, but some also from application fields. Table III and Table IV present information about the degrees and business activities of participants.

TABLE III. PARTICIPANTS ACCORDING TO THEIR BACKGROUNDS

Computer Science	Mathematics	Business Administration	Electrical Engineering	unknown
16	2	3	6	14

TABLE IV. PARTICIPANTS ACCORDING TO THE FIELDS OF BUSINESS ACTIVITIES

Programming / Software	Business / Data Analytics	Engineering / Development	Service	Unknown
17	2	7	1	14

As most of the participants had a professional relation to computers and computer sciences, most of them had programming skills in at least one programming language. A similar situation was seen for data analytics: only three of the

participants had no experience at the beginning of the course. All the others had worked at least with databases and SQL; many of them already had some knowledge on data mining and corresponding tools like R or KNIME.

In summary, the average of participants was on the novice level: they had more than a fundamental awareness; their focus was on developing a deeper knowledge and better skills in the area of machine learning and everything that belongs to it. But there were no potential users from an application area without any technical knowledge.

3) Student Participants

It was more challenging than expected to recruit students for participation. We observed a very high initial interest; however, only part of those students actually participated. Data about the reasons not to participate were not collected, but informal feedback indicates some plausible reasons:

- Conflicting time schedules with other compulsory courses

- The long duration of one year sometimes makes it difficult to fit to the personal study program; in particular, the master programs (with 90 ECTS credits) are designed with only one year for studies plus one semester for the final thesis.

- There was an uncertainty about expectations to the students for phase 2. As observed in other situations, students tend to go for the well-known, low-risk option.

Fig. 2 shows the number of students who successfully completed phase 1 of the project by study program. Various study programs have been grouped to clusters. The numbers clearly indicate that we mainly reached students of computer science and related subjects. There were no students from, say, mechanical engineering or biotechnology programs. This might be mainly due to organizational reasons, because the course did not really fit to their study programs. Eight students who gave up after the first week were not taken into account in Fig. 2.
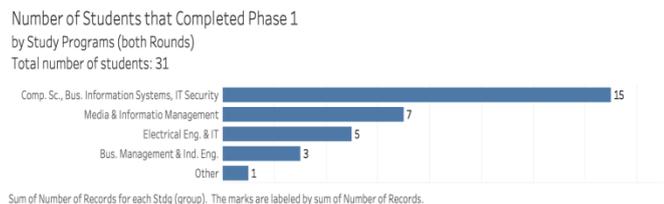

Number of Students that Completed Phase 1
by Study Programs (both Rounds)
Total number of students: 31

Sum of Number of Records for each Stdg (group). The marks are labeled by sum of Number of Records.

Fig. 2   Number of students in phase 1 by study program (only students who successfully took the oral exam were counted).

B. Phase 1: Practical Machine Learning

1) ML Course Design

In the student track, an interdisciplinary course "Machine Learning and Data Mining" consisting of lecture and laboratory/exercises was organized. The course comprised 4 hours per week. In this course, the students learned the same topics as the professionals in their track; however, the student track additionally included more advanced procedures and also focused more on theory. The students completed this phase with an examination (oral exam about a preselected topic) in accordance with the study and examination regulations and obtained the corresponding credits for this.

At the same time, the track for working professionals also qualified for the basics and procedures of ML. The qualification concept for professionals was based on a blended learning approach that combined the advantages of face-to-face sessions with digital, time and location independent forms of learning [8]. It consisted of a face-to-face kick-off event and a subsequent phase with interlinked online training sessions and face-to-face sessions as well as a closing event.

The online phases allowed for flexible knowledge acquisition of the participants with the inverted classroom principle [9]. The whole track was divided into different topics that were taught one by one during the track (cf. Table V). The various steps for each topic are explained in the following by the example of the 4th topic, classification. First, basics and principles of classification were presented and explained. Then several classification methods (decision tree, k-nearest neighbor, naive Bayes, random forest, …) were introduced in theory and with small examples. The final step was the evaluation of results. Self-assessments provided the opportunity for the participants to evaluate their understanding. In the following face-to-face session the whole topic (classification) was summed up and everybody could ask additional questions in the beginning. Subsequently, various classification methods were applied in a case study with the KNIME Analytics Platform [10].

Table V summarizes lectures and topics addressed in phase 1 for both tracks. As no prior knowledge about ML was assumed, basic topics like regression and classification also had to be part of the course. Due to time constraints, some further topics such as natural language processing were not taken into account.

TABLE V.   CONTENT UNITS OF PHASE 1

	Unit	Topics
1	Introduction to Machine Learning	The KNIME Analytics Platform, CRISP and KNIME, node repository, data manipulation, visualization
2	Explorative Data Analysis	data types, visualization, data description, statistical values, data preparation
3	Linear Regression and Basic Principles of ML	Supervised and unsupervised learning, linear regression with one and with several variables, overfitting, logistic regression
4	Classification	classification steps, decision trees, k-nearest neighbor, naive Bayes, random tree, model evaluation, confusion matrix, receiver operating characteristic
5	Deep Learning	principles of deep learning, neurons, perceptron, backpropagation networks, convolutional neural networks, deep learning environments and tools
6	Clustering	similarity and distance, k-Means and k-Medoid, hierarchical clustering, DBSCAN, validation of clusters
7	Association Analysis	Support, confidence and lift, apriori algorithm

2) Selecting the Tools: R versus KNIME

When doing or teaching ML in practice, there is always the question which platform to choose for implementation. There are numerous tools and programming languages to implement machine learning scenarios, many of them relevant in industry and academia. All tools follow either a traditional programming approach, so that students have to write code in order to implement a machine learning scenario; or a graphical paradigm: Implementing a machine learning example is done by "graphical programming", where the blocks of the program are represented by nodes with data input and data output that are connected drag and drop style.

The programming approach in ML is to a large extent dominated by the programming languages Python and R, both of them coming with a large number of libraries/packages related to machine learning and data science, and a strong support by the community. Beyond language and libraries/packages, there are additional tools such as RStudio [11] or Jupyter Notebooks [12] which support the development and deployment processes. Both Python and R are released under open source licenses. Due to its origin as a statistics package, R seems to be more focused on statistics and "classical" machine learning algorithms, whereas topics like Deep Learning are better represented in the Python world, mainly due to the popular TensorFlow framework [13].

The graphical approach is supported by various commercial and open source tools: For example, RapidMiner [14] and KNIME [10] are both desktop tools with additional server components for enterprise use; the Microsoft Azure Machine Learning Studio [15] is an example for a cloud based environment following the graphical approach. All mentioned graphical tools offer the execution of Python and R code as special steps.

For the present qualification program, we had to decide whether to go for the code based approach or for the graphical approach. The second decision was which tool or language to use. The criteria for the choice of tools were mainly two groups:

Which tool fits best for the heterogeneous group of participants, their backgrounds and previous knowledge (both students and professionals)? The tool should allow a steep learning curve, allowing the participants to focus on the machine learning problems and not so much on learning the tool/programming environment. Is there expertise with specific tools in the group of teachers providing the lectures?

Other requirements like performance or the availability of certain algorithms did not play a major role in the decision process. However, it was clear that the tool(s) to be used should be of practical relevance in industry, and there should be no licensing issues so that participants could easily install the necessary software on their own computers.

Since the course was open to students of all master degree programs, it was not possible to assume any programming knowledge, so it was clear that a graphical tool would be required in the student track. For the professionals track, we decided to involve the participants in the choice. Due to expertise of the instructors from previous courses and existing material, R was chosen as the option for programming and the KNIME Analytics platform as the graphical tool of choice.

As part of the first face-to-face session, a short introduction into KNIME was given. The practical part of the first unit, exploratory data analysis, was done with R and RStudio (after a short introduction into R). Instead of writing code from scratch, participants were provided with a framework program they had to complete. In the subsequent unit, linear regression, the exercises were done with KNIME. After this unit, feedback from the participants was collected informally. There was no clear opinion amongst the participants. The main arguments in favor of KNIME were:

- We can concentrate on machine learning instead of studying the syntax of a language

- Easier for beginners

The main arguments in favor of R were:

- As a software developer, it is easier to write code

- "Some things are more awkward with KNIME; in R it is just a line of code" (Quote form a participant)

- R seems to be more "professional", especially to software developers

Our observations confirmed that the group was quite heterogeneous in what they could achieve within the time of the face-to-face sessions, and many of the participants spent quite some resources in focusing on writing correct R code.

Finally, the decision was made to go for the graphical approach, using the KNIME Analytics Platform.

C. Phase 2: ML Projects

As introduced in Section II.C, the second phase of the qualification concept consists of real-world ML projects that enable participants to apply their knowledge from phase 1 to real-world problems.

The preparation for phase 2 already started towards the end of phase 1: Participating companies submitted project proposals that were checked for feasibility by the project team. Accepted proposals were then presented at the final face-to-face session (which was held jointly for students and professionals), and after the presentation each participant could name three projects (ordered by preference) that they would like to work on. The final assignment of participants to projects was done by the project team. Team size was between three and six people per project.

The practical phase started with a face-to-face kick-off meeting, in which the topics for the projects were presented in more detail by the representatives of the respective companies. At the end of the meeting, the project teams formed and team members could get to know each other (recall that students and working professionals had previously been in separate tracks). Also, first drafts for concrete steps in the projects were developed and agreed with the company representatives. For each team a supervising instructor or assistant was assigned. The development and outcome of the projects were measured in several plenary workshops with project presentations. Student participants also hat to write a project report about their work in the end.

A summary of all projects in both iterations of the course is given in Table VI. Titles are translated to English and anonymized by the authors. Four projects that started in iteration 1 were resumed in iteration 2.

TABLE VI. LIST OF ALL PROJECTS OF PHASE 2 IN BOTH ITERATIONS OF THE QUALIFICATION COURSE

Title	Area	Type of Problem	Platform
Improvement of customer conversion in an online office software	IT/CRM	Classification	Python
Ticket assignment and priorization	IT	Text classification	KNIME + Python
Predictive maintenance	Manufacturing	Classification, Time series analysis	KNIME
Framing of printed retail advertising	IT/Retail	Image recognition with deep learning	Python
Lead Classification – improve understanding of test users	IT/CRM	Classification	Python
Optimization of the factory acceptance test	Engineering	Classification	R
Classification of deviating production lots	Manufacturing of Semiconductors	Classification	R
Product recognition in the supermarket	IT/Retail	Image recognition with deep learning	Python + C++
Proposing customs tariff numbers for goods	IT	Classification, Text mining	KNIME + R
Predictive Maintenance	Engineering	Time series analysis	
Prediction of fine dust loads with Big Data	IT	Regression, Time series analysis	KNIME + R
Scheduling of machine jobs according to customer orders	Manufacturing	Optimization	Java
Anonymizing video recordings	IT	Image analysis with deep learning	Java, Python
Categorization of operating conditions	Engineering	Classification	
Correlations in test run data	Engineering	Time Series analysis	KNIME
Evaluation of Online License and Information Services	IT	Clustering, Association	Python + KNIME
Identifying related tickets	IT	Text mining	Python
Image text matching	IT	Image analysis	
Predictive forecasting of sales data	Business	Time series analysis	KNIME + Python + Tableau

IV. RESULTS AND DISCUSSION

A. Achieved Competencies

After phase 1, a survey was conducted to learn about the quality of the teaching and learning in that phase. Results can be seen in Table VII. Values could be chosen on a four-point Likert scale from 1 (do not agree at all) to 4 (fully agree). In the first iteration there were 16 returns from the 25 corporate participants, in the second iteration there were 10 returns.

TABLE VII. SURVEY AMONG PARTICIPANTS FROM COMPANIES ABOUT THE QUALITY OF TEACHING IN PHASE 1 (QUESTIONS TRANSLATED TO ENGLISH)

Question	Average round 1	Average round 2
The concept of education was learning enhancing.	3.25	3.5
The learning contents were conveyed intelligibly in the videos.	3.625	3.2
I was satisfied with the materials provided.	3.5	3.4
Through the learning videos, I was able to solve the tasks in the face-to-face event.	2.8	2.9

The participants rated the didactical concept quite good and it improved from iteration 1 to 2 (average 3.25 to 3.5). Those in iteration 1 especially liked the videos (average of 3.625) – despite the fact that these videos were just online-presentations based on slides. In both iterations, the participants were generally satisfied with the material provided (average of 3.5 and 3.4). Nevertheless the participants claimed that the videos were apparently not sufficient to solve the practical assignments in the face-to-face sessions (average 2.8 and 2.9, one of the lowest values we received). From personal feedback and from later surveys we learned that many participants would have liked to have more theory! This was astonishing for us, as we expected that theory, like the mathematical background of algorithms and quality of results, would be less interesting and not essential for professionals working in practice.

From the informal part of the evaluation, we list two quotes from participants that were quite typical (translated by the authors): "I have come to know the general machine learning process and the methods involved. I did not have any previous experience. In my opinion, I can go on working with this knowledge." And: "The theory part should be more beginner-friendly – more theory and more lessons."

B. Impact on companies

After the ML projects in phase 2 were finished we carried out another survey to learn more about those projects. From 12 participants out of 25 in iteration 1 we received the following results.

The question "Are you generally satisfied with the progress of your practice project?" was mostly answered very positive with an average of 3.94 on a scale from 1 to 5. The three participants who were not satisfied explained in a free text field that either time was too short to try out different approaches or that student members left the project team during the execution of the project.

Hence, we can conclude that the projects were very successful, but for upcoming iterations it will be helpful to bind the students more into the projects in phase 2 so that they will not leave prematurely.

It is important to emphasize how positive the feedback for the projects was, as only 62 % of the projects reached the goals they were heading for (question: "Have the defined goals of the project been achieved?"). All participants, without exception, claimed that they learned much about ML in the practical project. According to the background information given, the most important lesson learned was "data is king" – to have enough data, to have very good data quality and to make effort to prepare the data. We claim that you cannot teach this particularly "real-world" aspect of ML, learners must make their own experience with it.

One additional goal of the qualification project as a whole was to spread ML awareness. It seems that here we were especially successful in the participants' companies, as can be seen in Table VIII. It is also interesting to note that most of the practical projects were planned to be continued in the respective companies beyond the end of the qualification project (10 out of 12 participants).

Some other lessons learned in the ML projects in phase 2 are probably best described by the following quotes of participants: "The results were better than expected. However, more data is required. Also, it is necessary to 'play' with the data in order to continuously improve the results." And: "Training data must be chosen well."

C. Outlook

After two years of preparing and presenting contents, supervising ML projects and solving urgent ML problems, it was clear that the produced content and other materials should be kept and continued to be used. Hence, a comprehensive ML master module for continuing education was created, with the target group being professionals outside the university. It was carried through for the first time in fall 2019 with participants from different industries. Secondly, an online book on ML is currently being developed, based technically on the Bookdown R package [16] and augmented with exercises in R, Python and KNIME.

V. CONCLUSION

We have presented the concept and implementation of an integrated qualification program for machine learning targeted at graduate students of different major subjects and working professionals from industry, in particular SMEs. The focus is on practical aspects of machine learning; therefore the second half (one semester) of the course is spent working on real-world projects contributed by the participating companies. The completed course was held twice and evaluated each time. Overall, the program was received very well by participants. However, some findings from the feedback of participants suggest that teaching of theoretical concepts can be strengthened for the working professionals. Also, student dropout has turned out to be a problem that needs to be addressed in the future.

REFERENCES

[1] W. Sun and X. Gao, "The Construction of Undergraduate Machine Learning Course in the Artificial Intelligence Era," 13th International Conference on Computer Science & Education (ICCSE), Colombo, 2018, pp. 1–5.

[2] L. Huang and K. Ma, "Introducing Machine Learning to First-year Undergraduate Engineering Students Through an Authentic and Active Learning Labware," IEEE Frontiers in Education Conference (FIE), San Jose, CA, USA, 2018, pp. 1–4.

[3] S. D. Essinger and G. L. Rosen, "An introduction to machine learning for students in secondary education," Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), Sedona, AZ, 2011, pp. 243–248.

[4] B. Sakulkueakulsuk et al., "Kids making AI: Integrating Machine Learning, Gamification, and Social Context in STEM Education," IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), Wollongong, NSW, Australia 2018, pp. 1005–1010.

[5] E. Sulmont, E. Patitsas, and J. R. Cooperstock, "Can you teach me to machine learn? An exploration of pedagogical content knowledge for teaching machine learning to non-majors", In Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE'19), pp. 948–954, February 27–March 2, 2019, Minneapolis, MN. ACM, New York, NY.

[6] I. Döbel, M. Leis, M. Vogelsang, and H. Petzka, "Maschinelles Lernen – Eine Analyse zu Kompetenzen, Forschung und Anwendung," Fraunhofer-Gesellschaft, München 2018 (URL: https://www.bigdata.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer_Studie_ML_201809.pdf, last accessed 2020/04/09.

[7] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, London, UK: Springer-Verlag, 2000, pp. 29–39.

[8] C. J. Bonk and C. R. Graham. The handbook of blended learning: Global perspectives, local designs. John Wiley & Sons, 2012.

[9] M. J. Lage, G. J. Platt, and M. Treglia, "Inverting the classroom: A gateway to creating an inclusive learning environment," Journal of Economic Education, 31(1), pp. 30–43, 2000.

[10] KNIME Analytics Platform website. https://www.knime.com/knime-analytics-platform, last accessed 2020/04/10.

[11] RStudio website. https://rstudio.com/products/rstudio/. last accessed 2020/04/10.

[12] T. Kluyver, et al., "Jupyter Notebooks – a publishing format for reproducible computational workflows," in Positioning and Power in Academic Publishing: Players, Agents and Agendas (ELPUB), F. Loizides and B. Schmidt, Eds., pp. 87–90, May 2016.

[13] M. Abadi, et al., "TensorFlow: A system for large-scale machine learning." Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). 2016, pp.265–283.

[14] RapidMiner website. https://rapidminer.com/, last accessed 2020/04/10.

[15] Microsoft Azure Machine Learning Studio website, https://studio.azureml.net/, last accessed 2020/04/10.

[16] Bookdown website. https://bookdown.org/, last accessed 2020/04/10.

TABLE VIII. RESPONSES OF PARTICIPATING COMPANIES

Question	Average
Was the project also perceived by others in the company?	4.25
Does your company benefit from the project and its results?	4.06
Will ML continue to play a role in your company?	4.38
How relevant do you think ML is for your company?	4.5