

The QUA³CK Machine Learning Development Process and the Laboratory for Applied Machine Learning Approaches (LAMA)

Tuesday 22 October 2019 18:00 (20 minutes)

According to recent studies, Machine Learning has become a demanded skill in the world of engineering. Therefore, it is possible to find hundreds of online courses offering to teach these abilities. Some companies now call for 'Machine Learning Engineers' as a new hiring position. For many engineering students this is a very difficult situation, since taught courses are often focusing on the theoretical backgrounds of machine learning. However, in order to learn the correct use of machine learning methods, students have to apply them to real world problems. Hence, the scientists of the Institute for Information Processing Technologies (ITIV) at the Karlsruhe Institute of Technology (KIT, Germany) created the Laboratory for Applied Machine Learning Approaches (LAMA). We strive for a wide audience of engineering students, independently of their previous machine learning experience. The students go through three phases of learning. First, they get introduced to the theoretical concepts, then they will undertake guided hands-on experiments and lastly there is a creative 'Into-the-Wild'-part. This approach was very successful and was positively received by the students. The lab was designed around a novel machine learning process. This methodology is an adaptation of CRISP-DM, which is a unified cross-industry standard. However, it did not find general acceptance in academia. One reason might be the strong emphasis on business understanding.

We developed a simplified and easy-to-remember version of the CRISP-DM. This new machine learning process, which we named QUAAACK or QUA3CK, is an important part of the contribution of this work. It is an acronym for the individual process steps. The method commences with Question and Understanding the data. Followed by the iterative A³-process: Algorithm selection, data Adaption and parameter Adjustment. Based on these we will Conclude and compare the results and finally do the Knowledge transfer into a solution, product or publication. Figure 1 displays the process with precise description of each individual step.

! [Figure 1: The LAMA teaches the iterative QUAAACK process as part of the lab.] [1]

The laboratory is composed of two main parts: Guided assignments and the project based 'Into-the-Wild' phase. First, we teach students the basic concepts of frequently used machine-learning algorithms in theory and practice. We are using the widely adapted programming language 'python' for the guided part of the laboratory. Utilizing 'jupyter notebooks' we combine theoretical background documentation and its associated code in an interactive manner.

During the 'Into-the-Wild' part, students have the possibility to apply their learned skills to a real world engineering challenge. The aim of this creative part is to apply the QUA3CK process. Groups of two to four students can seek their own challenge or use provided datasets by the institute. In the end, all teams present their results to the other students and tutors. There was a broad variety of students' projects in summer 2019. Examples were a music composing neural network, the detection of sitting behaviour in order to improve posture and classifying whether bees are affected by toxins.

We can conclude that self-imposed goals lead to a better intrinsic motivation and a higher identification with the project.

Primary authors: Mr MEIER, Christopher (Karlsruhe Institute of Technology); Mr GRIMM, Daniel (Karlsruhe Institute of Technology); Mrs MOLINAR, Gabriela (Karlsruhe Institute of Technology); Prof. BECKER, Juergen (Karlsruhe Institute of Technology); Mr STANG, Marco (Karlsruhe Institute of Technology); Mr STOCK, Simon (Karlsruhe Institute of Technology); Mr HOTFILTER, Tim (Karlsruhe Institute of Technology); Prof. STORK, Wilhelm (Karlsruhe Institute of Technology)

Presenter: Mr STOCK, Simon (Karlsruhe Institute of Technology)

Session Classification: Submitted contributions

Track Classification: Artificial intelligence, data science, and machine learning