# Editorial

## Democratizing Machine Learning

Carsten Binnig

**Prof. Dr. Carsten Binnig**
**Professor for Data Management**
**TU Darmstadt**

Machine learning (ML) is today an important component of a large number of technical and social products. Whether we ask intelligent language assistants such as Apple's Siri on our smartphone for a kid-friendly restaurant with free parking, ask Google for a political fact check, or use autopilot functions in a vehicle, intelligent software is always behind it. The ever-increasing amount of data, which we collect about our world, has led to a paradigm shift: Instead of manually coding all the steps of knowledge processing, learning algorithms are used that allow systems to automatically recognize the structure of our world in data and to supplement their knowledge.

This paradigm shift is just beginning. On the one hand, in addition to user-generated data (business transactions, e-mails, pictures), the amount of data and the number of data sources in the area of machine-generated data are growing steadily. Experts predict that by the year 2020, 50 billion devices worldwide will be connected via the Internet of Things, which will generate continuous data that can be used to learn from. On the other hand, the spread of data-driven products is increasing rapidly. In addition to applications in the corporate sector (e.g., the prediction of failures of machines in production but also of complex autonomous systems), data-driven applications are already having an impact on the social or private sphere of individuals (e.g., healthcare). In the future, therefore, we will all have to interact, directly or indirectly, with intelligent systems.

However, there is already an alarming shortage of well-trained data scientists and ML experts, which is about to further increase in the future since more and more companies are moving towards using ML to solve a multitude of problems using data-driven methods. If learned systems are to be fully effective, then we have to make them much easier to build. One promising approach to overcoming this bottleneck is Automatic Machine Learning (AutoML), which simplifies the creation of data-driven products. However, existing AutoML methods are not suitable for use by domain experts. One reason for this is that AutoML methods do not cover the entire lifecycle for creating ML models, i.e., from data integration to actual model building to model evaluation.

Another important point, despite the problem of automating ML training is that decisions of trained ML algorithms differ from human thinking and their decision-making processes. This is essentially due to the different strengths of people compared to machines: While people often have little data (experience) but can rely on "common sense", ML algorithms score particularly with replicability and scalability with regard to complex but recurring tasks. Thus, a current direction is to investigate on how results of ML models can be explained to humans and how humans can interact as well as correct ML models.

Finally, for making ML usable in practice many other perspectives are important. For example, it is important to examine how to design ML models to increase the confidence and acceptance of users. In this context, an interesting question is whether models can capture the ethics of decisions and reflect our "gut feeling" for good and bad decisions. Other challenging questions are the economic aspects of ML and whether users are willing to pay for properties such as fairness or transparency.

The Data Science Institute will be a great catalyst to work on evaluating these new directions, to democratize ML, as well as to promote and exchange ideas between practitioners and researchers.