Peter Roßbach

Machine Learning, Modern Data Analytics and Artificial Intelligence – What's New?

Currently in the media, one may find numerous articles about the opportunities and future relevance of machine learning and artificial intelligence. Several success stories seem to prove this. As a consequence, universities and business schools are increasingly focusing on these topics. Questions, therefore, arise as to why these topics are currently in the spotlight, and whether they are just a hype or sustainable trends. The subject of this article is to explain what is new about the developments in these areas and why these developments have led to new potentials.

Big Data, Machine Learning, and Modern Data Analytics

Machine learning and artificial intelligence are not new, but there has been a great leap forward in recent years. Especially the advances in machine learning are the driving forces for modern data analytics and artificial intelligence. The aim in machine learning is to extract knowledge from data. Thus, it can be applied to nearly everything where data contains relevant information, from predicting customer needs and behavior to fraud detection, the prediction of capital markets, etc.

Going far beyond the abilities of traditional statistical analysis, modern data analytics reinvents data analysis by consequently utilizing the opportunities of today's digital technologies. There are many success stories about applying modern analytics; this leads to the question why modern data analysis is so successful. To answer this, we can identify four main reasons:

1. (Much) More Data for the Analysis

When I used econometric analysis for my research at the university in the mid 1980's, I had to type in the data manually from statistical yearbooks and other paper-based sources. Thus, the amount of data for the analysis was limited.

In contrast, nowadays we digitally record vast masses of data. Storage costs have decreased drastically. Every company can store the data which is generated in its systems in detail. Furthermore, it can purchase masses of data from external data providers to enrich their own data.

The data is very granular. Because of the progress in technology, there is no need to preprocess and aggregate the data anymore. In former times, with every aggregation, information loss was accompanied.

Moreover, we do not only record structured data, but we also collect unstructured data, created in social media, for example. Here we have sources for analyzing the attitudes and sentiment of customers for example.

Figure 1:
Traditional and
modern methods
[Sachan 2015]
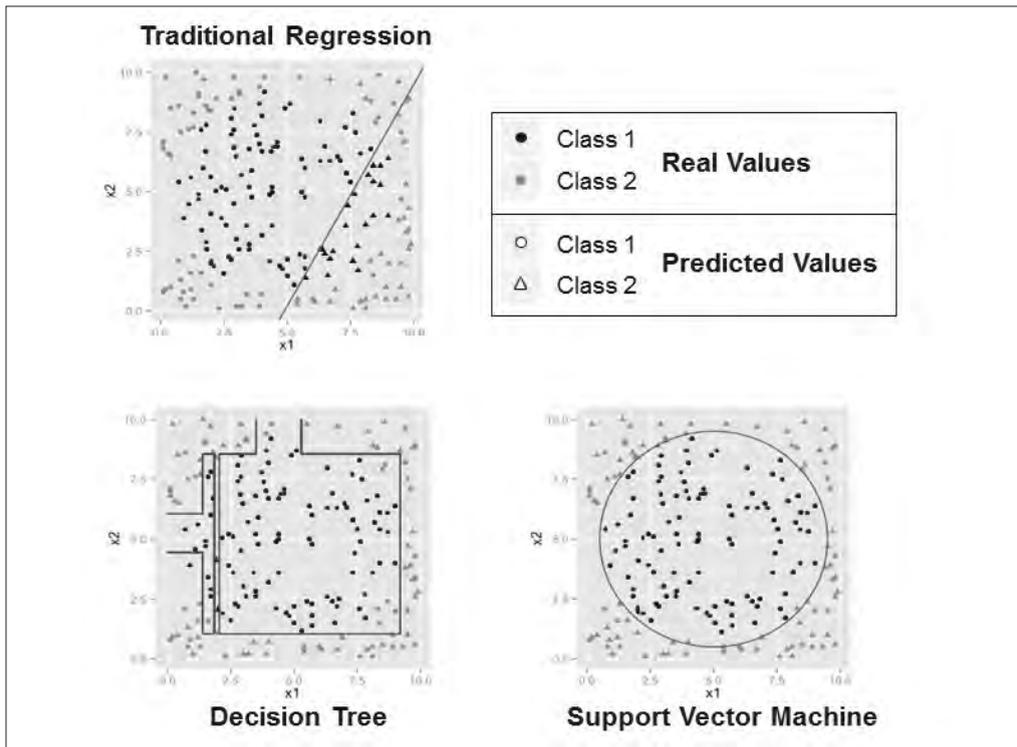
2. More Computing Power

The development of most of the classical statistical methods was influenced by limited calculation power. When Gauß developed the least squares regression method in 1795, only printed calcula-tion tables and simple mechanical calculation tools were available to support calculations. Even I learned statistics at school with a slide rule and later with a pocket calculator. As a consequence, the development of the methods was focused on performing as few calculations as possible resulting in strong assumptions like linearity or normal distribution.

Today, broad processing power is available, which continuously grows. Not only the processors of the computers are getting faster, but also special processing units designed for calculation are available, e. g. for neural networks or bitcoin mining. Furthermore, cloud services with a focus on calculation are available, which gives us the opportunity, to perform more complex calculations.

3. New Methods and Algorithms

Accordingly, a vast variety of methods and algorithms were developed in the past years, many of them under the term "machine learning". They are free of restricting assumptions like linearity or specific underlying distributions, and therefore can model the world like it is; nonlinear and dynamic. Figure 1 demonstrates the capabilities of these
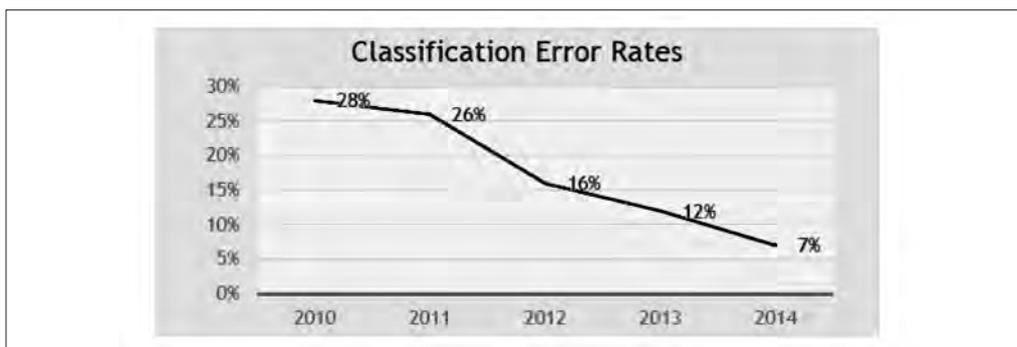
Figure 2:
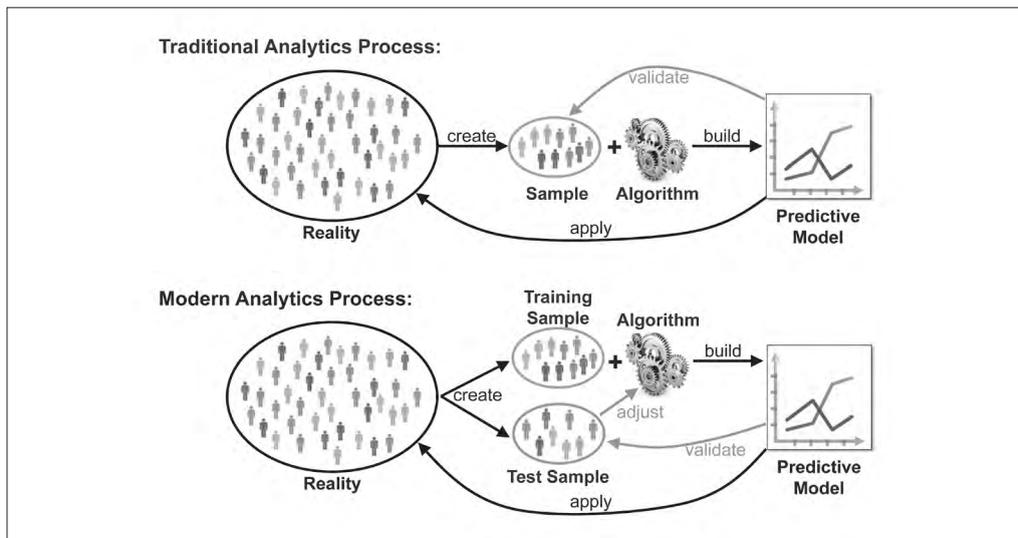Decrease of
Error Rates
[Brown 2015]

Figure 3:
Traditional Analytics Process and Modern Analytics Process

methods. The colors indicate the real class membership, e. g. creditworthy and not creditworthy or spam and not spam. The symbols, on the other hand, show the class membership estimated by the respective procedure. The greater the match, the better the method in this case.

A very promising family of algorithms is summarized under the term "deep learning". These are specific types of neural networks with specific learning capabilities. Their development caused a jump in the quality of data analytics in many areas. Figure 2 shows an example of the shift in quality when categorizing pictures.

4. New Analysis Processes

In traditional data analysis, a sample of the reality to be modeled is first generated (see Figure 3). In the next step an algorithm, like the ordinary least squares regression method (OLS), is used to build a model that best approximates this sample data. The quality of the approximation is then validated. If the quality is high, the model is applied to the reality, for example to make forecasts.

The underlying assumption here is that a high approximation quality will also produce a high forecasting quality. In practice this is very often not the case. Often the sample is not completely representative, and there are noise and outliers in the data. Usually, the algorithm is not able to distinguish between correct and incorrect data. This is called "overfitting" if the quality of the approximation (also called fitting) is high, but that of the application is poor.

In modern data analytics, the data is split into training and test data (see Figure 3). The training data is then used to build the model, and the test data is used to validate its prediction quality. Furthermore, in many of the modern algorithms, there exist parameters to adjust the sensitivity (also called complexity) of the algorithm against the data. The purpose is to filter out the irregular data and to concentrate on the underlying regularities. With this, the prediction quality can be maximized via the test data.

Figure 4 shows an example. In both graphs, the vertical axis represents the error rate and the horizontal axis the sensitivity of the algorithm. The left graph visualizes the approximation qualities for different sensitivities and the right graph the corresponding prediction qualities. The maximal approximation quality can be achieved with the maximum sensitivity. Here the model fits best to the training data. But if the model is applied to the test data, the result is different. For the test data, the quality increases when the sensitivity increases at first. But in contrast to the approximation quality, the prediction quality deteriorates after reaching a saddle

Figure 4:
Best Fit vs.
Best Prediction
[Roßbach/
Karlow n.d.]
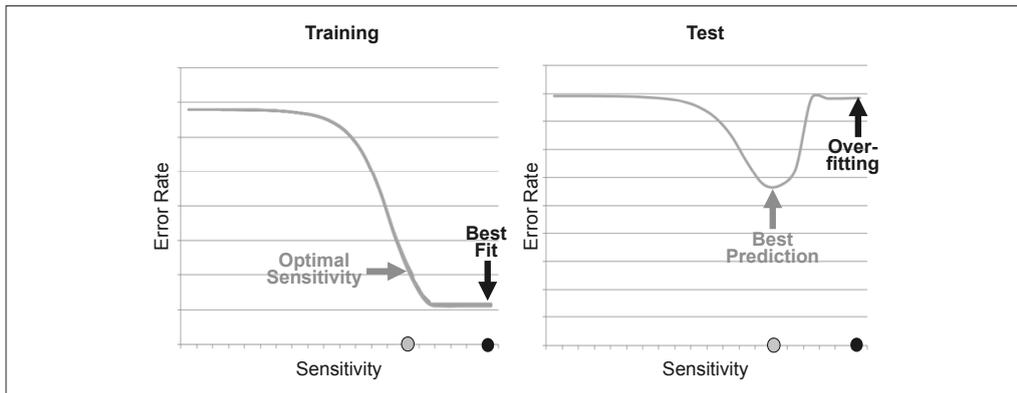
point due to overfitting effects. At this saddle point, the model produces the best forecasts.

The challenge is to find this optimal sensitivity, which cannot be computed directly. Thus, a massive number of calculations is needed to try different sensitivities. The aim is not to fit the model best to the data, but to find the underlying regularities.

Another technique is ensemble learning. Here, multiple models are created using different algorithms. The result is finally the combination of their predictions (see Figure 5). The models act like different experts having diverse domain experience. The purpose is to reduce the weaknesses and to combines the strengths of the algorithms used.

Artificial Intelligence (AI)

The progress in machine learning has led to a new generation of AI. AI is not a new research discip-

line. Researchers like Alan Turing and Joseph Weizenbaum studied how machines can mimic human intelligence in the 1950's. They thought about creating general problem solvers. We are still many years away from being able to develop those machines, if this will ever be the case.

When we talk about AI today, we usually mean domain-specific applications. The systems do not have general intelligence; they have a specific intelligence for a dedicated purpose, for example, for playing Go or understanding speech.

When I started to work in the area of AI in the early 1990's, we tried to transfer knowledge of human experts into systems. We used logic-based programming, e.g. in the form of rules and decision trees, to model the knowledge manually (see Figure 6). The challenge was to extract the knowledge from the human experts and to transform it into rule systems.
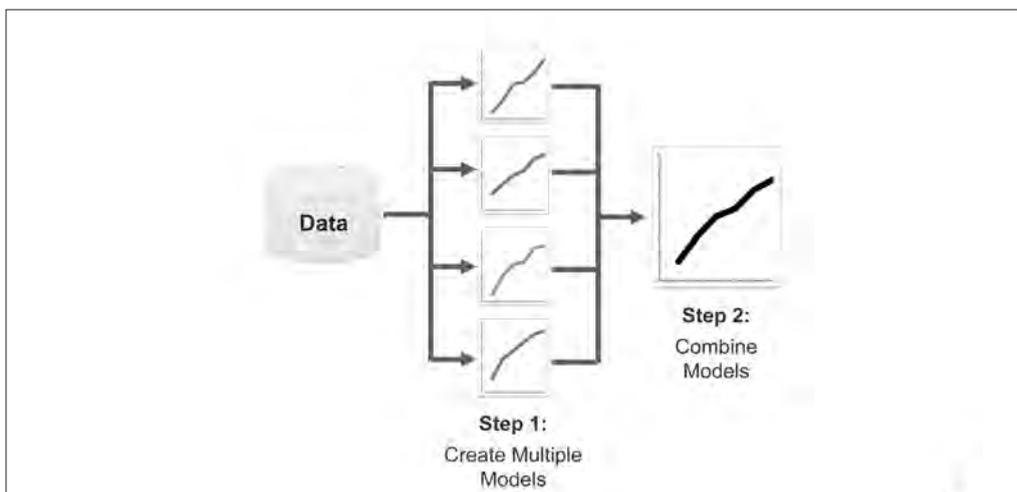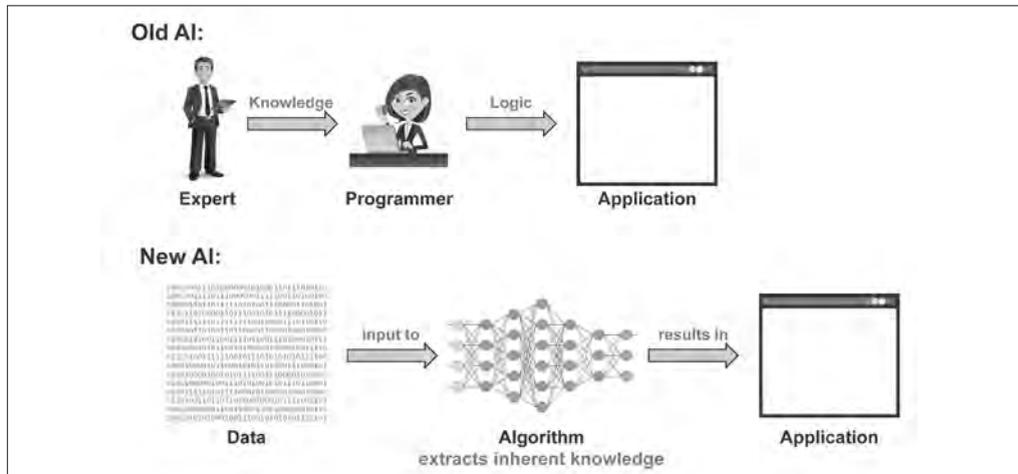
Figure 5:
Ensemble
Learning

Figure 6:
Former AI and
Self-Learning AI

We built advisory systems, e.g. for financial investments. The limitation was that the systems could never be more intelligent than the implemented knowledge. Furthermore, the systems often suffered from knowledge gaps. As a consequence, the first generation of AI systems was not successful.

Using machine learning, there is no need to program the knowledge manually anymore. Instead, the algorithms are fed with the data containing the relevant knowledge, and they acquire the inherent knowledge autonomously (see Figure 6). The developed systems, therefore, learn and adapt autonomously instead of simply executing predefined instructions.

The challenge is to provide the algorithms with the data that contains the relevant knowledge. This usually requires large amounts of granular data. Furthermore, the results have to be verified to ensure,

that the algorithms identified the regularities and not only spurious correlations.

A very promising technique is the so-called "reinforcement learning" which is a kind of smart self-learning. The algorithm trains itself by interacting with its environment. Based on a state, for example, a specific situation in a chess game or a position in a maze, it tries an action and gets a reward for the result (see Figure 7). The better the result, the higher the reward. Thus, it learns by itself via trial and error to perform the best actions. To learn strategies, the learning processes are not one-periodic, but multi-periodic. The higher the calculation capacity, the higher is the number of possible trials.

A very famous example is AlphaGo [Moyer 2016], the Go-playing system from Google. It was able to beat the world champions spontaneously. Go is one of the hardest games in the world for AI because

Figure 7:
Reinforcement
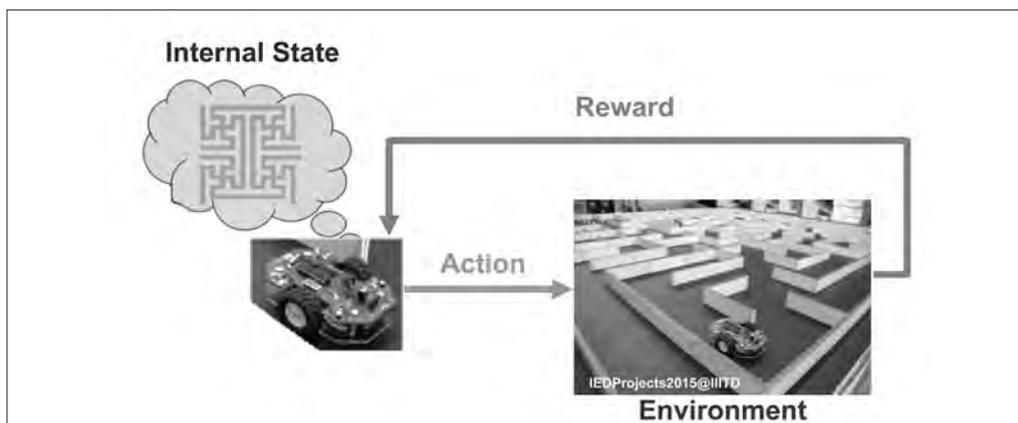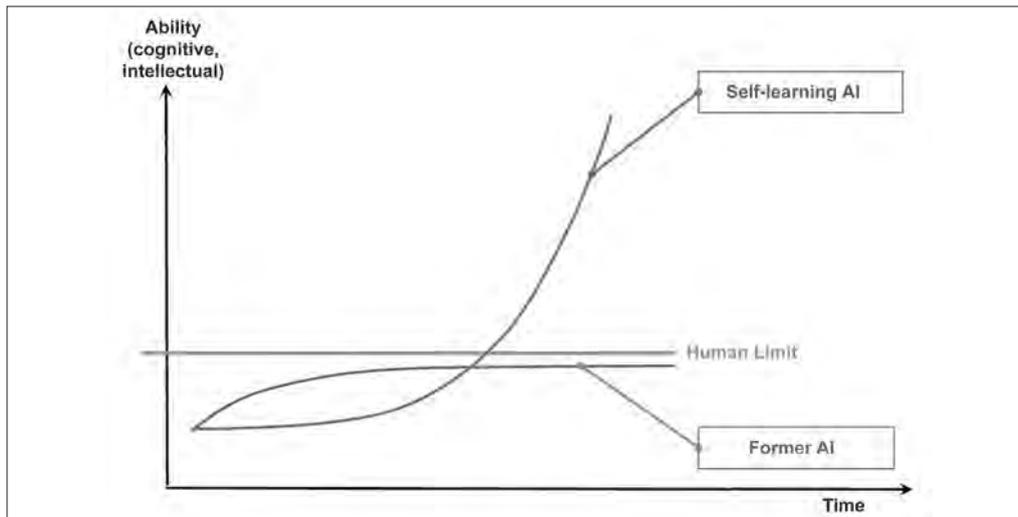Learning
(adapted from
[Shaikh 2016])

Figure 8:
Potentials of AI
(adapted from
[Furrer 2017])

of the huge number of different game scenarios and moves. The number of potential legal board positions is greater than the number of atoms in the universe.

The core of AlphaGo is a deep neural network. It initially learned to play using a database of around 30 million recorded historical moves. After this phase, the system was cloned and it was trained by playing large numbers of games against other instances of itself using reinforcement learning to improve its play. During this training, AlphaGo learned new strategies which were never played by humans.

Another milestone in the development of modern AI is Libratus [Solon 2017], an AI trained to play poker. In January, Libratus has beaten four of the world's best poker players in a 20-day tournament in a demoralizing way. Poker is a game with imperfect information, as many markets in the business context. Besides the ability to play poker, Libratus has to act intelligently in other ways. So it is required to bluff and correctly interpret misleading information.

In summary, it can be stated that the potentials of machine learning and artificial intelligence are enormous. The advances enable us to create applications that go beyond our human abilities (see Figure 8).

At present, the part of AI that is most being integrated into consumer and business applications is based on image processing, text processing, and speech processing. The quality of many applications based on these functionalities has reached 99 % accuracy due to the successes in machine learning. This is a critical threshold of making these applications acceptable for users and companies.

We are currently at the beginning of the development, and cannot yet foresee the range of possible applications. However, an ethical framework is needed to ensure that the technology is developed and used in the interests of people.

Sources

Brown, G. (2015). Deep learning for image classification, http://www.nvidia.com/content/events/geoInt2015/LBrown_DL_Image_Classification-GEOINT.pdf.

Furrer, F. J. (2017). Rezensionen, in: Informatik Spektrum, Heft 4, 2017, S. 393.

Moyer, C. (2016). How Google's AlphaGo Beat a Go World Champion, http://www.theatlantic.com/technology/archive/2016/03/the-invisible-opponent/475611.

Roßbach, P./Karlow, D. (n.d.). Structural Minimization of Tracking Error, Frankfurt School of Finance & Management (to be published).

Sachan, L. (2015). Logistic Regression Vs Decision Trees Vs SVM, http://www.edvancer.in/logistic-regression-vs-decision-trees-vs-svm-part1.

Shaikh, F. (2016). Getting ready for AI based gaming agents – Overview of Open Source Reinforcement Learning Platforms, http://www.analyticsvidhya.com/blog/2016/12/getting-ready-for-ai-based-gaming-agents-overview-of-open-source-reinforcement-learning-platforms.

Solon, O. (2017). Poker computer trounces humans in big step for AI, http://www.theguardian.com/technology/2017/jan/30/libratus-poker-artificial-intelligence-professional-human-players-competition.

Author
Prof. Dr. Peter Roßbach is Professor of General Business Administration, with emphasis on Applied Business Informatics and Information Technology at Frankfurt School of Finance & Management.

The article has already appeared as a blog post at http://blog.frankfurt-school.de/machine-learning-modern-data-analytics-artificial-intelligence.