



ALL OUR ACTIVITIES LEAVE A DIGITAL FOOTPRINT



text: Agnieszka Niewdana



Dr. Eng. Roman Simiński
Institute of Computer Science
Faculty of Science and Technology
University of Silesia
roman.siminski@us.edu.pl

Computers are omnipresent in our lives. They perform the necessary actions for us, capturing large amounts of information, which is now easy to obtain, process, and store. Information is recorded both by our computing devices by means of which it is saved on local data carriers (disks, flash drives) as well as in so-called computing clouds. Using computer systems and logging into the appropriate account and service allows us to perform personalized activities. We increasingly often use the option of saving our data “somewhere on the Internet.”

However, due to personalized resources and cloud-based services, information about us, our resources, and our activities is meticulously recorded by software managing the access to cloud computing and all other online services. Each of our activities leaves a digital footprint e.g. when we use social networks, online stores, or financial services. It is worth remembering that we also leave such a trace when making purchases in physical stores and using payment or loyalty cards. Actually, we do not leave it only when we pay cash.

Notwithstanding the fact whether the digital footprint makes it possible to identify us or we remain anonymous, it contains information about our activity e.g. what, when, and for what amount we purchased, what financial transaction we carried out, what we searched for on the Internet, where and how we traveled, what photos we took (when and where as well). Traces of our activity recorded in computer systems are associated with a number of significant problems related to, among others, confidentiality of information about us. Unauthorized use of this information for e.g. criminal purposes is both possible and dangerous. Information recorded by IT systems can, however, be used in many ways that are beneficial to us. Information recorded by the banking system over a certain period is not only a record of individual events, but also, in an implicit way, a reflection of the processes taking place in

terms of the funds in our account. A detailed analysis of a larger number of events from a certain period may allow for the deduction of many, often surprising facts about real events in our lives. For example, the end of a steady monthly income flow from a certain account may indicate a potential job loss, but the simultaneous start of a flow of income at a higher level may simply indicate that we have a better paid job.

The analysis of anonymous purchases in a self-service store can reveal an abundance of information important to the store's owner. The contents of the shopping baskets can enable them, for example, to select groups of goods which are most often bought together. Their identification can be used for such an arrangement and display of goods that the most frequently purchased ones are located in the immediate proximity.

These examples show simple applications of the currently very important field of artificial intelligence, namely intelligent data analysis, also known as data science. The concept of extracting knowledge from data is not new; it is derived from the concept of machine learning. The first known and successfully applied methods in

learn something new. To *learn* means to create a description containing previously non-existing knowledge about regularities, relationships, and trends found when learning data. The concept of machine learning was intended to *teach* the machine how to solve a problem in a way other than using the algorithmic method. The results of the algorithm can be different; they usually take the form of decision trees or decision rules. In both cases, the results of the machine learning algorithm allow to make an attempt to classify new cases.

Machine learning actually allows us to discover knowledge about the problem being solved. By broadening and generalizing this concept, we come to the concept of data mining. The goal of data mining is to discover previously unknown, useful, and non-trivial knowledge implicitly stored in data. Due to the fact that obtaining data for exploration may require additional activities, e.g. purification and preparation, and the result of the exploration – assessment and verification, a wider process, including exploration as one of its elements – is referred to as knowledge discovery in data. Nowadays, the ability to acquire data sets

The first and second decades of the 21st century were a time of great progress in the field of information technology. Computers have become faster, and their ability to store and process a wide variety of information is enormous. Every day we take pictures, have conversations via various communicators, send documents, or pay for purchases with a money card. We do not give any consideration to the question where this data is collected and how it can be used.

this field were developed in the second half of the last century. The best known algorithms are ID3, C4.5, C5.0 by Quinlan, AQ by Ryszard Michalski, a Pole living and researching in the USA. Machine learning algorithms are based on examples from which they are supposed to automatically

which may contain hidden but relevant knowledge is relatively easy and widespread, and these data sets can be large in terms of volume.