Machine learning is a process of data analysis whose goal is to understand the structure of a large amount of data. Using multiple algorithms, machine learning builds on a computer’s ability to find trends in and authenticate information. What separates machine learning from other methods of data analysis, is the fact that it automates the process of building data models. The computers are not programmed with specific instructions and are instead built to observe patterns in data that create the models. This is especially helpful when researchers are looking at data without any knowledge of trends within the population. Using machine learning, data scientists have the ability to parse through much more data in an efficient way.

The two main methods of machine learning are supervised learning and unsupervised learning. Supervised learning uses data in which the desired outcomes are known. The data scientists that enter the information tell the machine what the correct outputs are so that it can compare new data to that set. This type of learning is very helpful for research that seeks to classify data and make predictions. For example, a system that detects credit card fraud based on inputs by the customer.

Unsupervised learning takes a slightly different approach. While the machine is still fed large amounts of data, the “right answers” are not provided for it. The machine is programmed to find trends in the data without knowing any correct outputs. This type of machine learning is used with large amounts of transactional data where researchers have little background knowledge about the customer makeup. A classic application of this is when data analysts use clustering to identify customer segments based on a series of characteristics. This, in turn, can be used to recognize spending habits of consumers and help to inform a well-designed marketing strategy. Some data scientists use a mix of supervised and unsupervised learning depending on the type of data that they can work with. This “semi-supervised” method is ideal when the cost of using correct outputs for all the data is too high. Typically with this strategy, only a portion of the data shown to the machine will provide the desired outcomes with the majority of the date being raw.

Machine learning relies on algorithms in order to make something out of the data it is given. Some of the common algorithms that are used include decision trees, clustering, and neural networks. Decision trees are typically used when the classifications are already known and the model can predict the outcome of new data. It uses the patterns from previous outcomes to make decisions on future cases. Clustering is a process used with unsupervised learning. This means that the model takes all the data and uses it to group the data based on characteristics associated with it. The researchers can set however many clusters they want until the best model is reached. Neural networks are very advanced and attempt to mimic the biology of the human brain. Based on large amounts of data, the system is able to adapt and update its algorithms as it learns more.

There are many everyday applications of machine learning. One great example is the Facebook Newsfeed. If a Facebook user likes or comments on a friend’s post, the system will show you more content that was posted by that friend on your feed. Another example of this is the Netflix recommendation service. If you have a Netflix account and you watch certain types of movies, the algorithm will display related movies along with other commonly watched movies based on the behaviors of past viewers. As I mentioned earlier, fraud detection is a very common application of machine learning that is used by all major card carriers to ensure the protection of their customers. An application of machine learning that we might see more of in the future is the idea of self-driving cars. This is the ultimate use of machine learning, where dozens of factors must be considered. Especially for a task as complex as driving, the machine learning required for autonomous vehicles to be successful is going to be very high.

Going forward, we can expect to see many innovations within machine learning. This process has just recently begun to take off, with many companies taking advantage of the vast amount of transactional data available to them. Google has declared a primary focus on machine learning for themselves as a company. For over a decade, they have been offering courses to their engineers about machine learning and this dedication has been demonstrated through advancements in their algorithm. As more data becomes available, there will be more opportunity for this field of artificial intelligence to grow. It will be interesting where the technology goes next.

Sources:

<https://www.sas.com/en_us/insights/analytics/machine-learning.html>

<https://searchenterpriseai.techtarget.com/definition/machine-learning-ML>

<https://www.ml.cmu.edu/>

<https://www.wired.com/2016/06/how-google-is-remaking-itself-as-a-machine-learning-first-company/>