# Machine Learning

The field of Machine Learning contains concepts how computers can obtain information without explicitly programming this kind of information retrieval. These concepts of “Learning” can be divided in three main categories: Supervised learning, Unsupervised learning and Reinforcement Learning.

Supervised learning always deals with a user that ‘feeds’ input to the program as well as desired output. The program should recognize patterns that lead from the given input parameters to the desired output.

Unsupervised learning instead, is an approach where the program does have an input and needs to find a structure in those data. The finding of some pattern can be the goal of the program itself.

Using reinforcement learning, every output of the program is being valued by the user again. Output that has been found correctly will be strengthened, whereas incorrect values will act repulsive on the algorithm. After multiple iterations of this process, the program can find the best answer (but not always the correct answer) using the attracting and repulsive values.

Our goal is to use one machine learning technique in order to improve the outcome of our application. One major objective that can be addressed with Machine Learning is the relation between accounting record positions and how they are assigned to all the accounts that are important for this position.  
We identify two major problems regarding this classification:

1. What does a position represent?
2. Which accounts should be assigned to this position?

As we are processing an invoice, we will retrieve a position as a String. An accountant would be able to identify the position (which means a semantic identification of the object) and assign it to the accounts that are important in this matter. But, as there is no concrete rule which position belongs to which accounts, every company can apply this position to different accounts.

For instance, the maintenance of a car in the car pool of a company could be booked as car costs, or (if the company defines it more specifically) as maintenance, car parts and worker time.

Hence we need an algorithm that is capable of the following:

1. Assign involved accounts depending on the user (-> allow different account structure)
2. Learn relationships between a string and a set of accounts

While the algorithm should be able to deal with those problems, we will have another problem to deal with: OCR errors (e.g. “CAB” instead of “CAR”)[[1]](#footnote-1) and similar words (e.g. plural words such as “apples” instead of “apple”).

Keeping those constraints in mind, we can start thinking about a Machine Learning technique that satisfies our goal or at least helps us to reach it.

To narrow our search, we also have to think about automation. As this application should be able to reduce the time an accountant needs to process an invoice, we want to make this process as automatically as possible. Using a supervised machine learning method would lead to an application, that requires to validate each invoice every time. Hence supervised machine learning algorithms will not be considered here.

1. We used uppercase letters here to make the possibility of OCR errors between those two words more easily understandable. [↑](#footnote-ref-1)