



# Tackling document digitization through intelligent automation

How to exploit the potential of Machine Learning and Artificial Intelligence with the support of turicode AG

turicode AG  
Technoparkstrasse 2  
CH-8406 Winterthur

[info@turicode.com](mailto:info@turicode.com)  
[www.turicode.ch](http://www.turicode.ch)

# Intelligent Automation

Everything you need to know about Machine Learning in digitization processes

## Introduction into the world of intelligent automation

As digital transformation is becoming an integral part of today's society, a growing movement towards digital automation solutions is taking place. The global market is expected to reach €10.7 billion by 2023.<sup>1</sup> Document digitization, meaning the transformation of human-readable documents into a digital, machine-readable form is among the most important factors driving digitization in general, and it is a fundamental pre-requisite for automated text and data analytics. For any company, document digitization represents a first step in the automation of their business processes and their roadmap to digital transformation.

### But why is an automated software solution necessary at all?

Employees around the globe are currently executing these every-day tasks predominantly manually. Copy-pasting is still the most common technique to extract valuable information from documents reason being is the complexity in dealing with multiple document types, layouts, structure and context of use. This manual work is costly, repetitive and error prone. By digitizing information-intensive processes, costs can be cut by up to 90%, improving turnaround times by several orders of magnitude, as well as enhancing the overall operational efficiency of businesses.

Over the last few years, many companies have been relying on new technologies such as Robotic Process Automation (RPA) or Optical Character Recognition (OCR) to both simplify and accelerate manual work steps. This enabled, for example, the formatting of data in a spreadsheet to be fully automated through the use of RPA technologies. However, in practice, it became evident that the area of application of such automation technologies is limited to smaller data volumes and structured information.<sup>2</sup>

It is estimated that only 20% of data is already accessible in a structured format, while 80% of relevant corporate data is unstructured and "locked" in PDFs, Word files or emails, and therefore unsuitable for further analysis or integration with digital services.

### Hidden potential and benefits

This is the moment when "intelligent automation" comes into play. Thanks to the integration of artificial intelligence in the automation processes, it is now possible to cope with larger and complex unstructured data volumes.<sup>2</sup> Using semantic knowledge and methods of artificial intelligence (AI), such as machine learning (ML), deep learning (DL) and natural language processing (NLP), organizations can automate manual data transformation in a comprehensive way rather than punctually. Consequently, the systems are able to understand the content of documents as humans would do.

#### Definitions<sup>3</sup>

##### *Artificial Intelligence (AI)*

AI refers to technologies that simulate work processes that normally require human intelligence. These include meaning extraction from texts and images, the formulation of predictions or recommendations and the recognition of patterns.

##### *Machine Learning (ML)*

Algorithms that learn and can be trained to predict or make decisions based on a specific set of data.

##### *Deep Learning (DL)*

DL is a specific variant of ML, which combines several layers of model parameters in the learning process. Two representative examples are speech recognition and image recognition.

##### *Natural Language Processing (NLP)*

Innovative technology solution, which is able to recognize or reproduce human language. Examples are chatbots or automatic analyses and processings of customer mails.

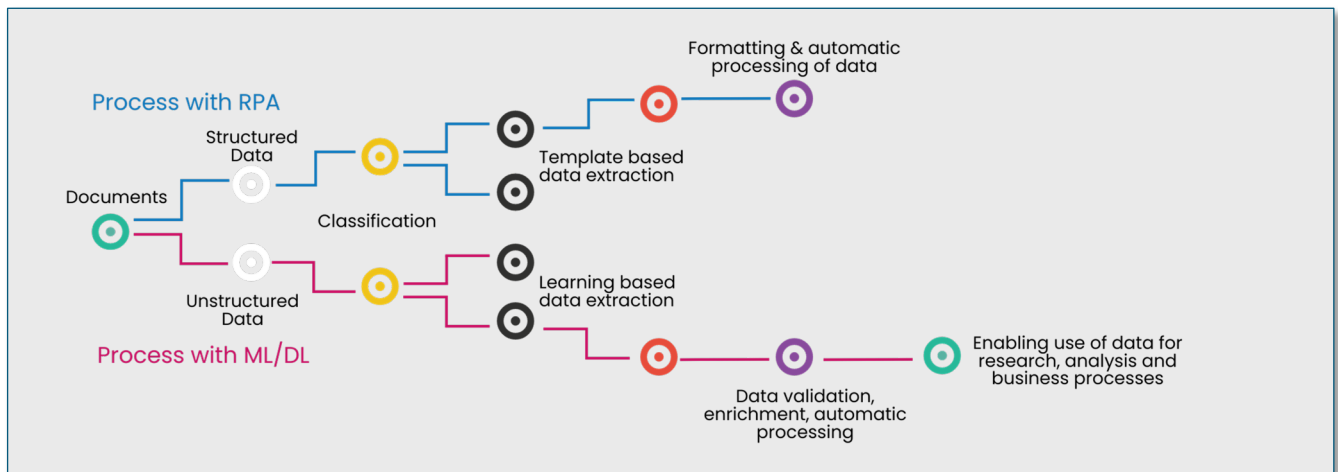


Figure 1: Comparison automation process with RPA vs. ML/DL

The major strength of this type of intelligent automation lies primarily in the numerous measurable benefits that can be achieved within a very short time period. Among the most prominent advantages are without a doubt the extensive cost savings as well as time savings. However these are not the only exciting potentials that are hidden behind all of this:



Costs & Revenue  
Cost reduction of about 50% and simultaneous revenue increase



Time Savings  
90% less time spent on document processing



No programming skills  
Business users can steer the entire process without programming skills



Working 24/7  
Fully automated service – Working 24/7 – Reduced error quote



No Layout Standard  
Works with a variety of document layouts - regardless of language

## Overcoming skepticism about innovative tech-solutions

There are still many companies that are hesitant to implement new AI technologies in their businesses. The reasons for this are very diverse. However, the most common barriers include lack of budget, concerns about data protection, application to isolated use cases, fear about insufficient inhouse tech knowledge and skepticism about the accuracy of the data output.

Many of these fears have been at least partially mitigated during the last years, given the technological process and the overall change of behavior towards these kind of tools. The corporate world is becoming increasingly accustomed to the innovative solutions and at the same time the algorithms continue to advance. There are already elaborated systems that enable companies to use ready-to-use SaaS applications with no high upfront investments. Furthermore, with the option to choose between an on-premise solution or a cloud model, companies can adopt their preferred service to secure their data privacy. The flexibility of the machine learning models allows companies to set up training and learning models without technical expertise or external dependency. Thanks to this user-friendliness and the possibility of deploying the systems directly into existing IT infrastructures, Machine Learning as well as Deep Learning solutions help companies to process diverse data both quickly and cost-effectively.

To demonstrate the accuracy of predictions made by deep learning systems in comparison to human predictions, turicode conducts studies with customer data sets using its in-house developed software MINT.extract. Thereby, the accuracy of the predictions in the training are set

in relation to the Inter Annotator Agreement (IAA). The IAA value stands for the human agreement / consistency evaluated by labelling the same sample documents twice by two different annotators. For instance, the IAA value for the project was 0.945. This is a kind of upper limit that predicts how good an optimal machine prediction could theoretically be with this training data.

The precision of the predictions are illustrated in the following figure:

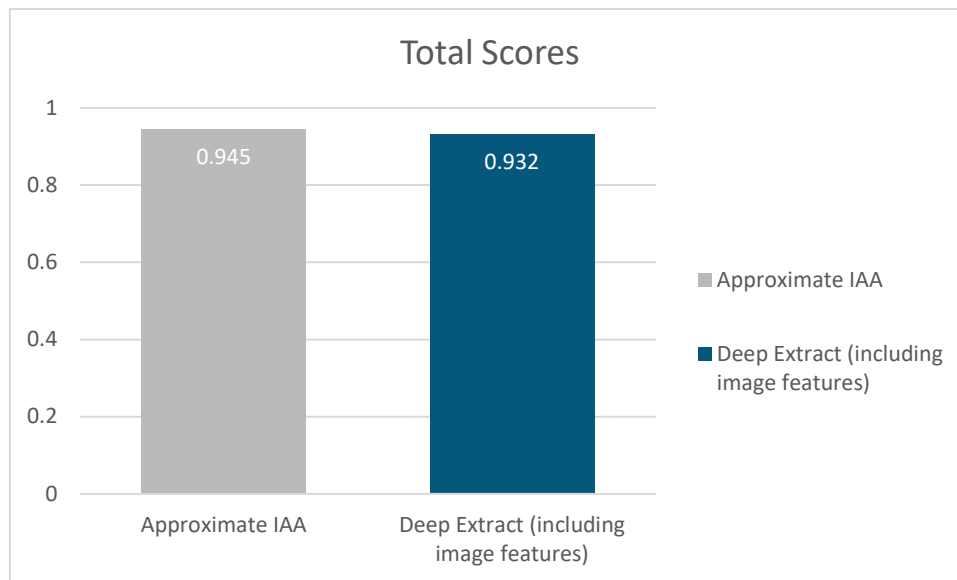


Figure 2: Comparison of Machine vs. Human Prediction

## Conclusion

With Machine Learning and Deep Learning techniques, intelligent document automation solutions are about to launch a new era of information retrieval engines for documents. Disruptive software solutions are not limited anymore to specific document types or require technical skills for setting up own models. Therefore, business users remain in the driver seat of their individual digitization processes by steering and enhancing the intelligent models over time. This way, document processing is applied where the expert knowledge is. All in all, there are still challenges ahead in the field of Artificial Intelligence and Machine Learning. The path to a generic AI solution, which can handle a wide range of individual use cases, is not entirely paved yet, but different companies are already working on its development. The question to what extent machine and deep learning as well as intelligent automation will change in the near future remains unanswered. But one thing is already clear - it promises to be exciting.

Any Questions left? – Contact us!



**turicode AG**  
*info@turicode.com*  
*www.turicode.com*  
*+41 52 520 62 02*

## References

- <sup>1</sup> CISION PR Newswire (2018). [The growing need to optimize business processes for meeting the dynamic requirements of customers is expected to drive the overall digital process automation market.](#)
- <sup>2</sup> Swiss Post Solutions. (2020). [Intelligente Automatisierung auf der Überholspur zur gesteigerten Kundenzufriedenheit.](#)
- <sup>3</sup> Deloitte (2019). [Automation with intelligence. Reimagining the organisation in the 'Age of With'. Deloitte Insights.](#)