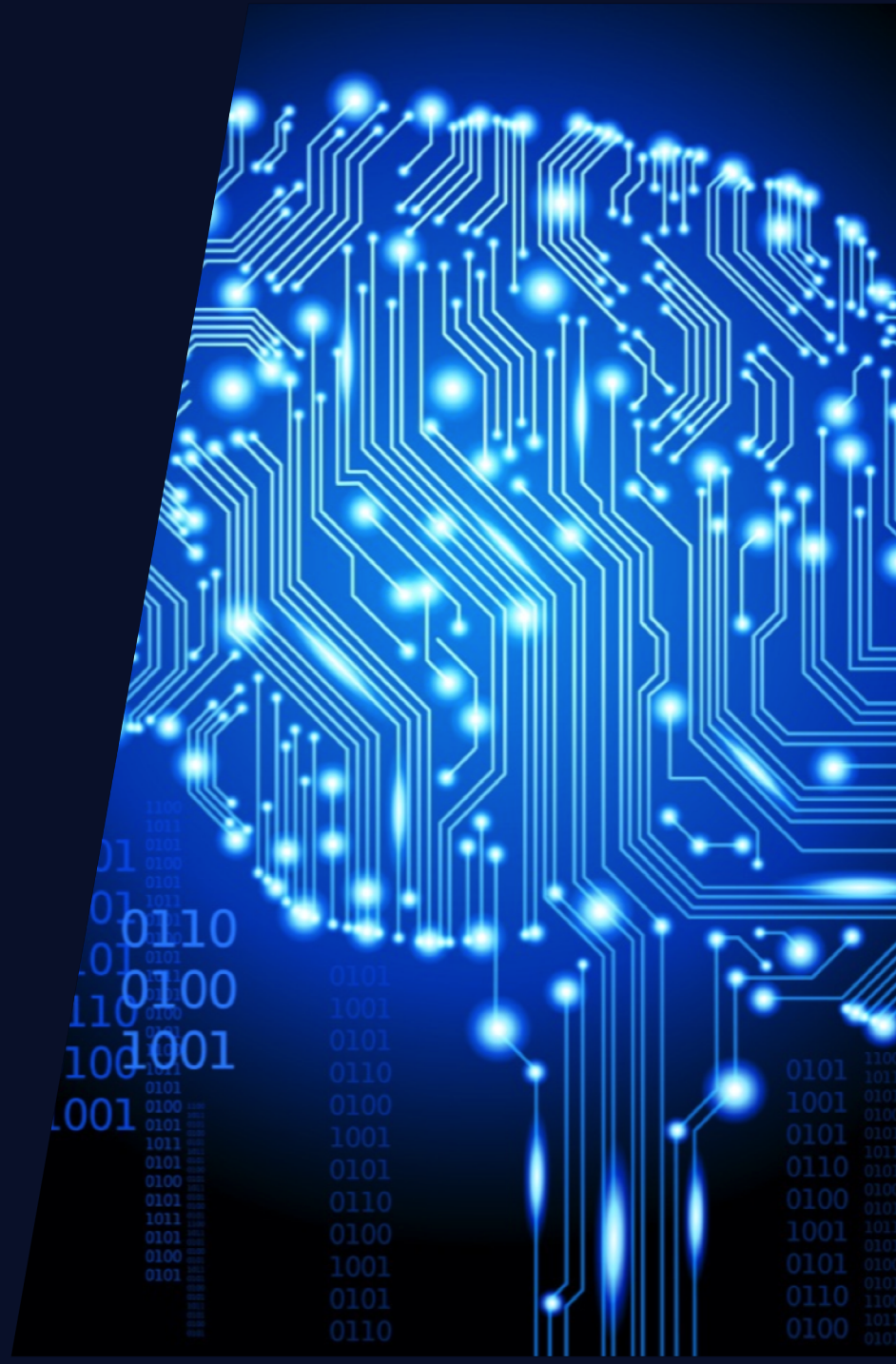
doculayer®

Interactive Machine Learning

Doculayer - Cognitieve Content Management



Human – AI interaction

Introduction

- AI is a powerful tool that can achieve high performance when **clean data** is provided.
- Compared to academic datasets, **real world datasets** often **are** ill-defined and **noisy**, thus academically proven AI algorithms might struggle to provide high quality output.
- **Human feedback** is crucial in any AI system and helps to **improve** the Machine Learner when it faces difficulties.
- Our integrated AI system monitors itsself for weaknesses (low confidence) and ask the user for feedback *when appropriate*.

Human – AI interaction

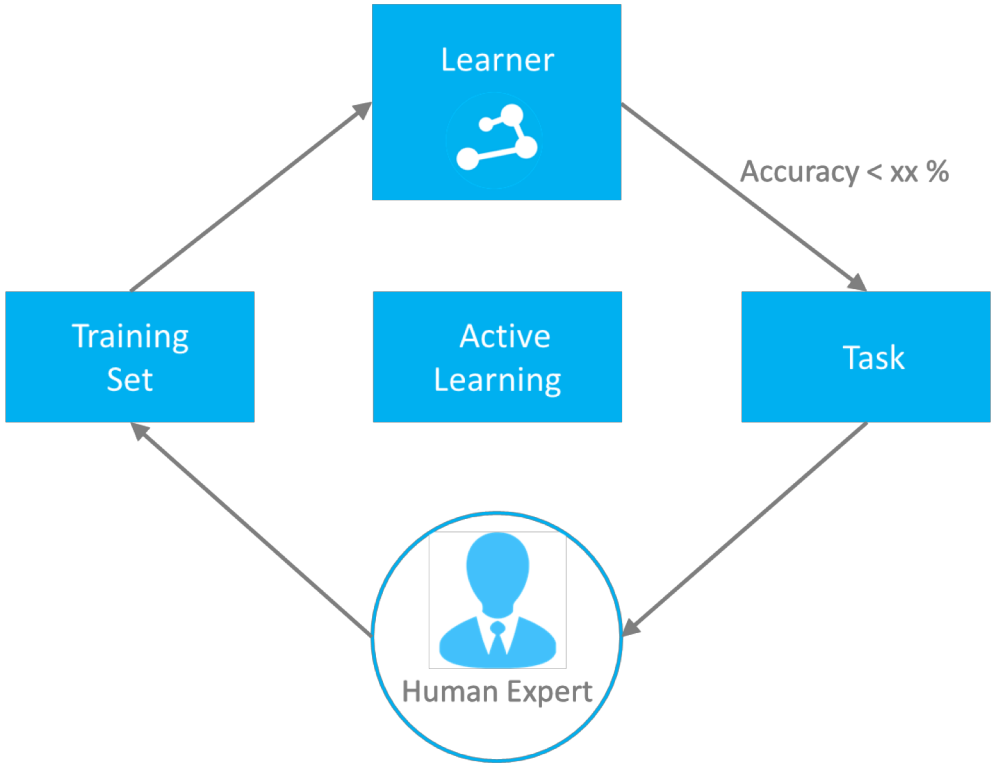
How does machine learn?

Active Machine Learning
and
Interactive Machine Learning

Active Learning

Machine is in charge

- System periodically sends out workflow check-tasks to users to measure classifier performance.
- Tasks are made for items with the lowest confidence in the decision and the highest expected performance gain.
- User classifies manually when performance drops below threshold, system takes back control when performance is above threshold.
- Corrections are added to a training set and the classifier is retrained.



Interactive Learning

Synergy between man and machine

- Active Learning only allows the Machine Learner to initiate a task for correction.
- However, in interactive learning also the user can give advice to the Machine Learner when he spots a mistake.
- Interactive learning combines the best of both worlds.

Interactive Learning

Case 1: Classification (metadata)

- Doculayer ML classifies content items into predefined classes (e.g. based on document types) and fills in the corresponding metadata field.
- If verification of the generated metadata is needed the Machine Learner creates a workflow task.
- The user verifies the metadata values marked for verification and if needed applies corrections.
- Training set is updated and model is retrained.

Interactive Learning

Case 2: Clustering

- Clustering: grouping similar text documents & images/videos together based on attributes such as extracted keywords.
- Easily extendable to a classifier.
- Hierarchical clustering orders the groups top-down allowing efficient data exploration.
- For clusters with the largest information gain, Interactive Learning generates tasks for a user to review.
- The Machine Learner improves based on the field expert input and produces a better representation of the data.

Interactive Learning

Case 3: Named Entity Recognition (NER)

- Named Entity Recognition (NER) is the process of identifying specific groups of words which share common semantic characteristics (e.g. company names, dates, serial or social security numbers).
- Doculayer ML identifies entities within a text document.
- When the system has a low confidence the Machine Learner creates a task.
- The user corrects the output if needed by adding missing entities, unmarking falsely identified entities or providing context.
- Corrections are added to the training set and trigger a retraining of the model.

Thank you for your attention!

About Doculayer

Doculayer is a Cognitive Content Management platform with Smart Content Solutions for contracts, employee files, marketing assets, and archiving.

For more information, visit https://www.doculayer.com

Prinses Catharina-Amaliastraat 5
2496 XD The Hague
The Netherlands
+31 (0)85 303 76 47
info@doculayer.com

Meeuwenlaan 100 (Pand Noord)
1021 JL Amsterdam
The Netherlands

doculayer

Find the Real Value of your Business Content.

www.doculayer.com