Machine Learning in Legal Tech

Degree programme : BSc in Computer Science | Specialisation : Data Engineering
Thesis advisor : Prof. Dr. Erik Graf

The field of legal tec has seen increasing automation in recent years. Companies are introducing contract management solutions to have all contracts located in one single tool. Most of these tools are supporting their users with automations that are developed as traditional software not using ML/AI. This work develops guidelines for the evaluation of ML potential and a decision framework to answer the question if ML is ready to be used in legal tech.

Concept

In a field that is in completely hyped in media, it is not advisable to directly address ML/AI providers with a lot of cash to let them "create your solution". Since this paper will not deliver the methods how to explain to your superiors that investments into new technology resulted into money being spent without results, we will not use this first approach. This thesis focused on the considerations that must be considered when evaluating ML/AI solutions. In order to find out if solutions can be delivered for your task you need to evaluate your tasks in a structured way. This way is using resources and will cost you time and money, but it is the only way you can increase the certainty to get a working solution. While finding solutions to integrate into contract management is the main driver of this thesis, it will show you all steps you need to make before you can start developing your interface. The structured test method of this thesis will assist you to find the solution to which integrating is worthwhile. After using the structured evaluation approach often only a PoC can confirm whether if a solution really is fitting your needs. Based on those steps we want to develop a Cookbook that delivers a list of items to consider when evaluating and a Checklist with which you can confirm what type of solution can be applied for your needs.

Goals

This bachelor project is a single project that focuses on defining the steps to evaluate multiple solutions. The target is to evaluate possible data sources for defining a test set and document the main issues that you need to tend to. Additionally, it is laying out the special considerations that need to be taken into account in the field and explains the need to invest time into the creation of good test data and structured testing methods. Those methods are then applied during:

– Search for training material
– Annotation of training material
– Search and evaluation of ML solutions
– Performing PoC
– Calcuation of their performance
– Decision making if solutions are delivering as promised

Results

The test set was successfully created during the course of the project. It was then used to evaluate the potential of ML/AI in the complex field of contract management. The results of one solution against the annotated test set was exceptionally good. The system "contract intelligence built by cortical.io" is made to be used by contract managers not having a data science degree. It was able to perform extractions with high certainty after only 40 of 200 contracts have been annotated to create a model.

It uses a convenient way to display the predictions of the AI that is shown in the screenshot below. The predictions are then confirmed by humans and used again to improve future models and their predictions. Additionally, to the test set, a catalog of questions has been laid out that serves as guideline to test ML Solutions in the field of legal tec. The catalog can also assist at the considerations that need to be considered when one wants to create a proper ML Solution in this field.

Roland Angelo Roccaro
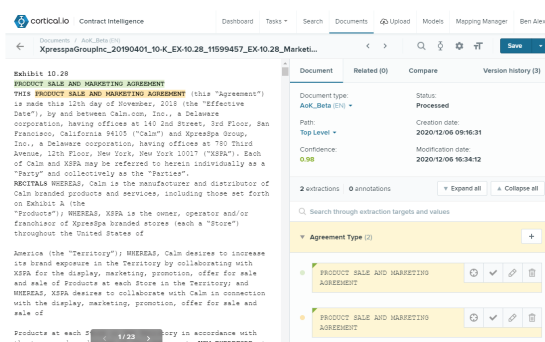
Contract Intelligence of cortical.io , extraction of agreement type