

Machine Learning

– Extracting Data from Bond Documents

Executive Summary

Extracting key data from long, unstructured legal documents has long been left to manual admin effort, or rudimentary extraction tools with cumbersome rulesets and low success rates. Intelligent Document Processing is changing this and making it possible to extract data accurately and consistently.

The client currently spend a significant amount of manual effort extracting key data from Bond and Loan documentation. Intelligent Document Processing is fast becoming an invaluable tool for many businesses in this area, so they are naturally very interested to understand what is possible and how it could benefit them.

About the Client

The client is the leading and most experienced provider of bond trustee and loan agency service. We have over 3,000 active assignments in the non-bank lending sector for more than 850 issuers/ lenders from 30 countries.

They are also a provider of high-quality bond market data through its subsidiaries Stamdata and Bond Pricing. Data includes detailed reference data, price and index information on all debt securities issued by the public sector, financial institutions and corporates.

Since October 2021, the client has been part of Ocorian group. Ocorian is a global leader in corporate and fiduciary services, fund administration and capital markets. It has USD 270bn in assets under administration and employs more than 1,350 professionals in 20 offices spread across Americas, EMEA, and Asia.”

The Challenges

Key bond data such as the return type, first payment date, interest rate/spread, covenants etc can be written into bond documents in a myriad of wordings and structures. This makes traditional extraction tools hard to program, as paragraphs of text are used instead of structured tables, and various wordings for the same things can be used across large documents leading to expanding and reprogramming extraction rules for every eventuality.



CASE STUDY

This is especially apparent when part of your business model is structuring and providing this data to the open market. Extracting efficiently and consistently then becomes especially important: the more you can extract, the more you can sell.

The Solution

The client engaged psKINETIC as a first step to understanding the potential of using intelligent document processing for their use case. A proof of concept was developed in Google Cloud Platform using Auto ML. The model was trained using a set of ~150 example bond documents: a mix of High Yield and Investment Grade notes, with both floating and fixed rates.

The model was trained to extract 7 common fields of varying types, and the performance of the model was then assessed. The results confirmed that even with the minimum training requirements for a model of this type, the extraction performance metrics (Precision & Recall) were suggesting that the use of this technology would be appropriate for the use case.

Results

Playing back this proof of concept to this client opened their eyes to the possibility of using intelligent tools from the large cloud providers such as Google, and how this could transform their extraction process with minimal custom development. Although a custom model may come into its own in the future, a public cloud pre-built model, trained on the correct documents, can yield positive results.

This approach will move the business away from the endless development of manual rules currently required to maintain their extraction tool, and towards the constantly evolving and flexible role an intelligent model can play in the process.



psKINETIC

psKINETIC is a leading provider of Intelligent Automation Solutions for Financial Services.

Our focus is on delivering outcomes and financial return for clients at pace. We don't believe in 'just rip out and replace' – it is too risky and too costly – with our Intelligent Glue we help clients connect and enhance existing systems and processes. Our experienced professionals master a range of automation technologies and we support customers through the full cycle of Design, Build, Run & Optimise.

For more information, visit www.pskineti.com