

MORTGAGE AND LOANS

Intelligent Data Extraction and Automation



SCALE

The following points help to emphasize the scale of the challenge with mortgage documentation:

- over 600 document types
- most types completely unstructured (i.e. no set layout) or with a highly variable structure
- over 250 document types requiring some level of data extraction
- on average about 8 individual data elements required from each document processed.

INDUSTRY

Mortgage and Loans processing

PRODUCTS

Aluma

THEMES

Document Classification
Data Extraction

CHALLENGE

Our client is a regulatory technology provider focused on developing innovations that enhance the transparency and accuracy of the mortgage process and improve the quality of loans. By early 2018, they had established themselves as a leading provider of technology and services for the mortgage industry, and as a result of their dominant position, they were already processing over 120 million documents a year on behalf of their customers, and with those numbers set to rise steeply.

SOLUTION

In an initial meeting with Aluma, we explained how our innovative approach to document automation, through a unique blend of machine-learning, NLP and rules-based techniques, could offer something qualitatively different from the other products they'd trialed in terms of approach, and also quantitatively different in terms of automation levels, accuracy and speed of implementation. Interested to investigate further, our client made available a set of a few thousand mortgage documents comprising hundreds of different types for an initial benchmark. 80% of the documents were used to train Aluma's AI classification engine and the remainder were retained as an independent test set. The training and test process took just a few minutes, was entirely automatic and required no manual adjustment of parameters.

The results at this early stage were already impressive – the detailed statistics created by the test process showed >70% automation at >99% accuracy, meaning that more than 70 in every 100 documents could be confidently classified, and of those less than 1 in 100 were incorrectly assigned.

The AI engine was even able to point out that some of the training samples that had been incorrectly labelled! Once these had been corrected and the training was enhanced with additional samples of the less common types, the learnt model was at a standard ready for production use. The entire process was performed by our client themselves.

IMPACT



CLASSIFICATION

The simple training procedure outlined above has now produced automation levels for document classification of well over 80% with very high accuracy (99.5%+).



EXTRACTION

Similarly, the data extraction is able to pull out over 88% of all required data elements from the documents. Both of these have dramatically cut operational costs.



COST REDUCTION

“It was straightforward to integrate Aluma's technology into our AWS cloud architecture – the documentation and support was excellent – we were amazed to find we could classify up to 15 documents per second in a single service, so were able to keep hosting costs to a minimum.” Director of Technology.